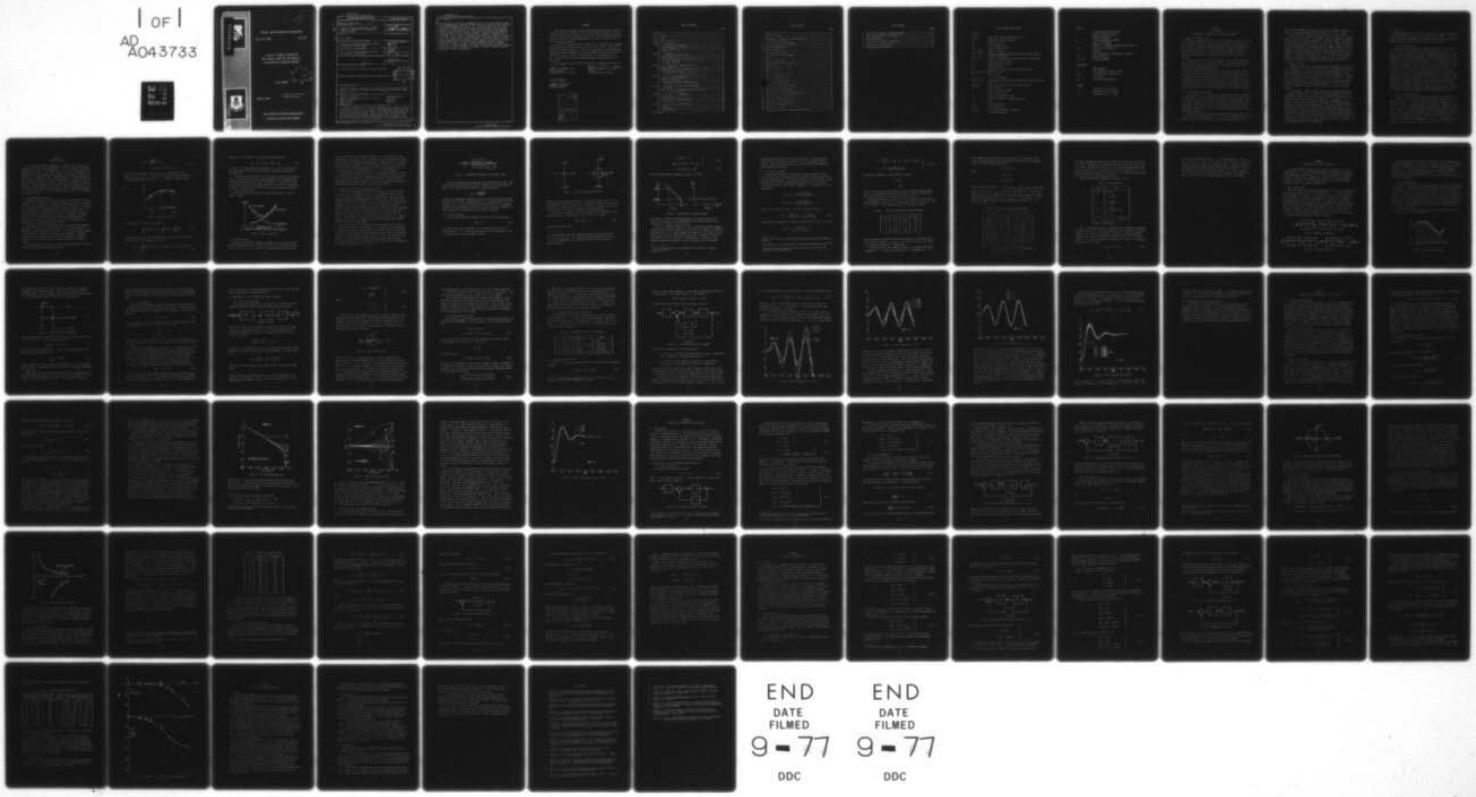
AD-A043 733 FRANK J SEILER RESEARCH LAB UNITED STATES AIR FORCE --ETC F/G 12/1
A STUDY OF TUNABLE INTEGRATION AND CONTROL THEORY FOR THE ANALY--ETC(U)
MAY 77 M L SABIN

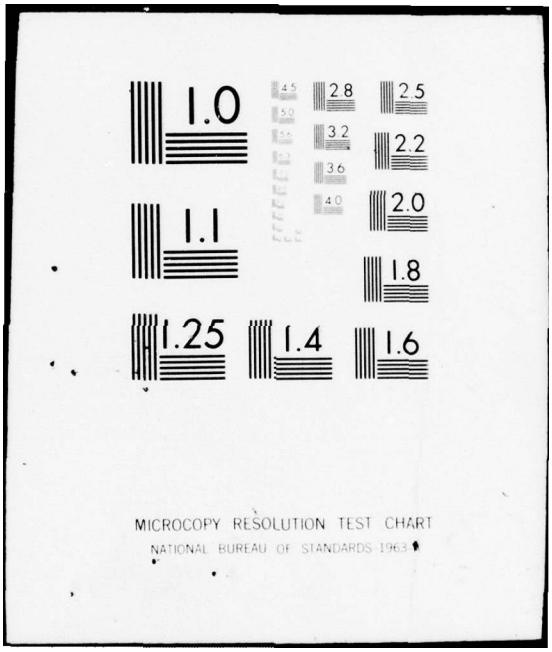
UNCLASSIFIED FJSRL-TR-77-0006 NL

1 OF 1
AD A043733



END
DATE FILMED
9-77
DDC

END
DATE FILMED
9-77
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

FRANK J. SEILER RESEARCH LABORATORY

SRL-TR-77-0006 MAY 1977

A STUDY OF TUNABLE INTEGRATION AND CONTROL THEORY FOR THE ANALYSIS OF DIFFERENTIAL EQUATION SOLVERS

FINAL REPORT

PROJECT 2304

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

AIR FORCE SYSTEMS COMMAND

UNITED STATES AIR FORCE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER FJSRL-TR-77-0006 AD A-	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* A STUDY OF TUNABLE INTEGRATION AND CONTROL THEORY FOR THE ANALYSIS OF DIFFERENTIAL EQUATION SOLVERS		5. TYPE OF REPORT & PERIOD COVERED Final Report January 75 — December 76 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Marc L. Sabin		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Frank J. Seiler Research Laboratory (AFSC) USAF Academy, Colorado 80840		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS DRS 61102F 2304-F1-55
11. CONTROLLING OFFICE NAME AND ADDRESS Frank J. Seiler Research Laboratory (AFSC) USAF Academy, Colorado 80840		12. REPORT DATE May 1977 13. NUMBER OF PAGES 72
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* 16 2304 17 F1		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Contains information presented at the Symposium on Air Force Applications of Modern Control Theory, July 1974 and at the Modeling and Simulation Conference, April 1975.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Numerical Integration
Numerical Solution of Differential Equations
Tunable Integration
Control Theory
Error Analysis
Frequency Response
Root Locus
Bode Plots
Stability

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Tunable integration is an approach to the numerical solution of ordinary differential equations that is conceived, developed, and employed based upon the principles of linear control theory. This present work complements the original publications on tunable integration by Smith, and it also reiterates and extends the earlier work by this author. The properties of the zero-order-hold tunable integrator and the control-theory methods used to analyze those properties are equally important in the context of this report.

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 319204B

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT

Beginning with a discussion of numerical error from a control-theory perspective, ~~we present~~ the concept of an ideal integrator having no error but that due to finite computer word length. The lack of adaptability of the ideal integrator motivates the tunable integrator, which possesses the requisite flexibility. The majority of the report covers the analytic development of tunable integration, the frequency response of the zero-order-hold tunable integrator, and a root-locus analysis of that integrator as employed in a first-order, linear system. Throughout the presentation, emphasis is placed on the use of these control-theory techniques to determine stability and to tune the integrator for improved performance. The final section of the report stands somewhat apart from the rest, but is equally important in that it discusses the effects of computer-code implementation upon the frequency response of a software package. The results of the studies reported on herein are only a start. The potential promised by tunable integration deserves and requires further investigation.

FOREWORD

This report was prepared by the Applied Mathematics Division, Directorate of Aerospace-Mechanics Sciences, Frank J. Seiler Research Laboratory, United States Air Force Academy, Colorado. The work was initiated under Project Work Unit No. 2304-F1-55, "Tunable Integration." Major Marc L. Sabin was the Project Engineer-in-Charge.

I would like to give particular thanks to Mr. J. M. Smith, whose comments and counsel were invaluable during the course of this effort. I would also like to give thanks to Capt R. P. Fuchs of the Department of Astronautics and Computer Science, USAF Academy, for his many suggestions during the research and for his critical review of the manuscript. My sincere thanks go to Mrs. D. Weiss for her efforts in typing the manuscript and preparing it for publication.

This technical report has been reviewed and is approved for publication.

MARC L. SABIN, Major, USAF
Deputy Director
Aerospace-Mechanics Sciences

BURTON H. HOLADAY, Lt Col, USAF
Director
Aerospace-Mechanics Sciences

FOR THE COMMANDER

THOMAS M. TOMASKOVIC
Chief Scientist

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ACRONYMS

English

Arg {*}	Phase angle of the complex quantity {*}
C(s)	Compensator transfer function
E	Total local error
e_R	Local roundoff error
e_T	Local truncation error
$f_x = \partial f/\partial x$	Partial derivative
$G(j\omega)$, $G^*(j\omega)$	Frequency-response and sampled frequency-response transfer functions, respectively
G(s), G(z)	Laplace and sampled-data transfer functions, respectively
GH	Open-loop transfer function
H	Feedback transfer function
K	Forward path gain
M	Magnitude of frequency-response transfer function
$M_{db} = 20 \log_{10} M$	Magnitude in db
$p = \lambda T\gamma$	ZOH-TI gain
$q = (\gamma-1)/\gamma$	ZOH-TI zero location
$R_c(s)$, $R_i(s)$	Compensation and integration reconstructor transfer functions, respectively
$R_{ZOH}(s)$	Zero-order hold reconstructor transfer function
$r = \omega/\omega_s$	Sampling ratio
s	Laplace transform variable
T	Integration step size
T_{CL}	Closed-loop transfer function
TI	Tunable integration or tunable integrator
t	Time
u(t)	Unit step function
Z{*}	Z-transform of {*}
ZOH-TI	Zero-order-hold tunable integrator
$z = e^{sT}$	Z-transform variable

Greek

γ	TI phase-compensation parameter
Δ	Forward difference operator
ζ	Damping coefficient
λ	TI gain-compensation parameter
τ	System time constant
ϕ	Argument of frequency-response transfer function
$\psi(z)$	Characteristic polynomial
ω	Frequency
ω_d	Design frequency or damped natural frequency
ω_n	Natural frequency
$\omega_s = 2\pi/T$	Sampling frequency

Subscripts

c	Computed value
k	kth increment or value at t = kT
mm	Mixed-Mode-Euler computed value
t	True value
te	True-Euler computed value

Other

$\in$	Inclusion; e.g., a $\in$ [b,c]
$\lvert * \rvert$	Magnitude; e.g., M = $\lvert G(j\omega) \rvert$
$\dot{*}$	Derivative; e.g., $\dot{y} = dy/dt$

CHAPTER 1

Introduction

"The Purpose of Computing is Insight, Not Numbers"[1]

1.1 The Purpose of This Report

The principal objective of this report is to generate some of the insight that Dr. Hamming succinctly stated to be the purpose of all computation. Specifically, I will present a novel technique of numerical integration for the solution of ordinary, first-order differential equations occurring in initial-value problems; and I will analyze that method from a control-theoretic viewpoint that is applicable to the analysis of any numerical algorithm.

Intended readers include control engineers, who can greatly extend the efforts discussed herein, and numerical analysts, who can use these concepts along with their own effective methods to further enhance our ability to numerically solve differential equations.

A third audience to whom this report is addressed is comprised of the faculty and cadets of the United States Air Force Academy who are involved in teaching and studying numerical methods. Specific applicability is intended for courses in the Departments of Astronautics and Computer Science and of Mathematics. To be useful to these people, the style of this report is tutorial, and much is said that will be obvious and elementary to the reader experienced in the fields of control theory and/or numerical analysis.

1.2 The Importance of the Subject

The United States Air Force could not perform its mission, it could not "fly and fight," if it did not possess the means to numerically compute the solutions to differential equations. The motions of a hand grenade thrown by a soldier, a bullet fired from a rifle, an aircraft on a strafing pass, a smart-bomb following a laser designator, a ballistic missile on an inter-continental trajectory, or a satellite orbiting the earth are all governed by differential equations.

Differential equations describe more than the motion of objects. The Riccati equation for optimal control is a differential equation, equations of chemical kinetics are differential equations, the covariance matrix of a

Kalman filter propagates according to a differential equation, and electric circuits are described by differential equations; the list is endless.

In most realistic situations, analytic solutions cannot be found. The need for numerical solutions is self-evident: they are needed to accomplish operational missions, they are needed for simulation (especially important with the increasing amount of training via simulation), and they are needed for design and analysis. Different situations, however, require different approaches to their solutions. The formalized differential-equation-solver packages such as DVDO, GEAR, DE[2], etc., are excellent for highly accurate solutions on large, ground-based systems. For the small airborne computer, simpler, less generalized approaches are required, such as the technique employed in the ballistic trajectory algorithm of Duke, et. al.[3]

1.3 Tunable Integration

A number of years ago, J. M. Smith thought it reasonable that since the digital computer is a discrete operating device, the techniques of sampled-data system analysis and digital filtering could effectively be used to design methods for numerical integration on the digital computer. Thus was born the concept of tunable integration as it is presented herein. Smith has used the technique with considerable success.[4] Only little, however, has been formally written concerning the method[5],[6],[7],[8], and this report is intended to increase its familiarity within the Air Force community.

The technique is based upon the synthesis of a discrete approximation to continuous integration. The result is a difference equation in which adjustable parameters provide the capability to directly control the characteristics (pole location, frequency response) of the integrator. This capability makes tunable integration a technique that proffers tremendous benefits in terms of ease and flexibility of use, large regions of stability, controllable accuracy (via means other than variable order or step size), and simplicity. High-order integrators of this type have been developed by Smith[4],[5], but this author feels that more significant benefits accrue to the Air Force from the use of the low-order form in appropriate problems such as airborne fire control. Thus, this report will restrict its attention to the low-order form of tunable integration.

1.4 Overview

The organization of this report is based on using tunable integration as a vehicle for demonstrating the application of control theory in analyzing methods of numerical integration. The integration algorithm itself and the method of analysis are equally important parts of this report.

1.4.1 A Brief Outline

Chapter 2 presents a general discussion of errors and how to define error criteria in terms of the pole location and frequency response. The concept of "ideal" integrators is used to demonstrate these error criteria and to motivate the tunable integrator. Though the development also appears elsewhere[5],[7], the complete formulation of the zero-order-hold tunable integrator is presented in detail in Chapter 3. An example problem is solved using that integrator and clearly demonstrates the tuning property. Chapter 4 presents the frequency-response characteristics of the integrator. In Chapter 5 a root-locus analysis is performed and stability is evaluated. Also in that chapter is a short demonstration of how the frequency-response and root-locus analyses can be used to initially tune an integrator. The importance of program implementation is discussed in Chapter 6, wherein the difference in properties between two program codes using the same basic method, but programmed differently, is shown. Shampine, et. al.,[2] have noted the significance of differing implementation upon the results obtained with GEAR and DIFSUB, two packages using the same fundamental methods.

1.4.2 Background Knowledge

In order to understand this report, the reader should be familiar with some basic concepts of linear and sampled-data control systems, and of numerical analysis. Specifically, the reader should be somewhat familiar with the following: Laplace transforms, block diagrams, transfer functions, root-locus analysis, frequency-response analysis, control-system stability, phase and gain compensation, sampling of continuous signals, the sampling theorem, Z-transformations, difference equations, polynomial interpolation, and classical methods of numerical integration. The true integrator is continuous and the numerical integrator is discrete. Thus, familiarity with concepts in both continuous and discrete control theory is important.

CHAPTER 2
A Question of Errors

In his "Essay on Numerical Methods"[1], Hamming cites five main ideas germane to the study of numerical methods. The first of these is quoted at the opening of Chapter 1. Two other ideas, or concepts, are truncation and roundoff error, the fundamental error sources of any numerical method to be implemented on a digital machine. Perhaps the most important consideration for any numerical metnod is its error characteristics. The methods of control theory provide a different perspective on the errors inherent in a numerical algorithm. Tunable integration is a method of numerical integration that is designed to take advantage of that perspective and to thereby provide excellent error characteristics. In this chapter we shall look at the fundamental question of errors.

2.1 Some Elementary Preliminaries

Roundoff error arises from the finite length of the "word" in a digital computer. Hence, any number that cannot be represented by a finite number of binary bits (most numbers) must be rounded off by the computer. The problem is unavoidable, and, as Hamming suggests, one can only do one's best to minimize its effects. A principal problem with roundoff is that it tends to worsen as the number of computations increase or as computer word length decreases. Thus, roundoff can be especially severe when using a mini or micro computer to solve a differential equation over a relatively large (compared to the integration step size) interval.

Truncation error, on the other hand, tends to require smaller integration steps, since it arises from the necessity to use finite polynomial expansions to represent arbitrary functions. In a typical classical approach to the development of numerical integrators, the integrand is expanded polynomially and a truncated series from that polynomial is analytically integrated over a specified interval to provide the desired integration formula.

As an example, let us derive Simpson's integration formula.[9] Newton's forward interpolation formula can be written as

$$\dot{y}_k \approx p_k = \sum_{i=0}^{n} \binom{k}{i} \Delta^i \dot{y}_0 \qquad n \leq k$$

$$= \dot{y}_0 + k\Delta\dot{y}_0 + \tfrac{1}{2}\, k(k-1)\Delta^2\dot{y}_0 + \ldots \tag{2-1}$$

where p_k is the polynomial expansion of $\dot{y}$ to n forward differences, $\Delta\dot{y}_m = \dot{y}_{m+1} - \dot{y}_m$. Let n = 2 so that we are approximating $\dot{y}$ by a quadratic in k over the three point interval $\dot{y}_0$, $\dot{y}_1$, $\dot{y}_2$ as shown in Figure 1 below.

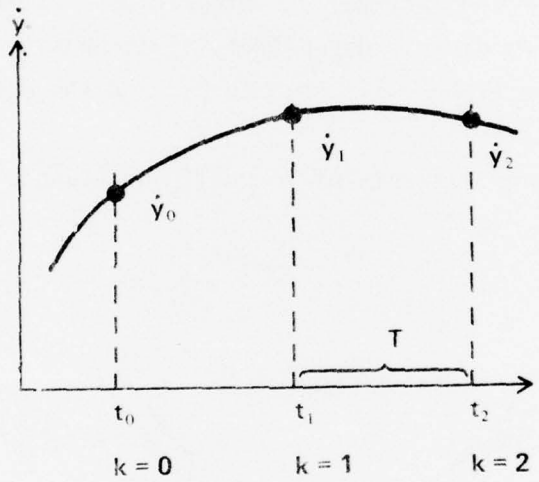


Figure 1. Quadratic Fit to $\dot{y}$

The integration over the interval gives us

$$\int_{t_0}^{t_2} \dot{y}(t)dt = y_2 - y_0 \approx \int_{t_0}^{t_2} p_k dt = T\int_{k=0}^{2} p_k dk \tag{2-2}$$

where we note that dt = Tdk. Substituting for p_k from Eq (2-1) and then expanding the forward differences, we obtain

$$y_2 \doteq y_0 + T\int_{k=0}^{2} [\dot{y}_0 + k\Delta\dot{y}_0 + \tfrac{1}{2}\, k(k-1)\Delta^2 y_0]\, dk = y_0 + \frac{T}{3}\,(\dot{y}_0 + 4\dot{y}_1 + \dot{y}_2) \tag{2-3}$$

Equation (2-3) is Simpson's rule, which has the general form

$$y_{n+2} = y_n + \frac{T}{3}(\dot{y}_{n+2} + 4\dot{y}_{n+1} + \dot{y}_n) \qquad (2\text{-}4)$$

To obtain it, we approximated the integrand $\dot{y}$ by a polynomial truncated to the second order (in k) and then analytically integrated to obtain the desired result.

Truncation error arises from all of the neglected terms in the expansion of p_k. It represents the error in approximating the true integrand by the selected polynomial over the interval of integration. Clearly, the larger the interval over which a given order polynomial is made to fit an arbitrary function, the larger the errors will tend to be, and the more significant will be the truncated terms.

The contradictory requirements of roundoff and truncation error are shown in Figure 2.

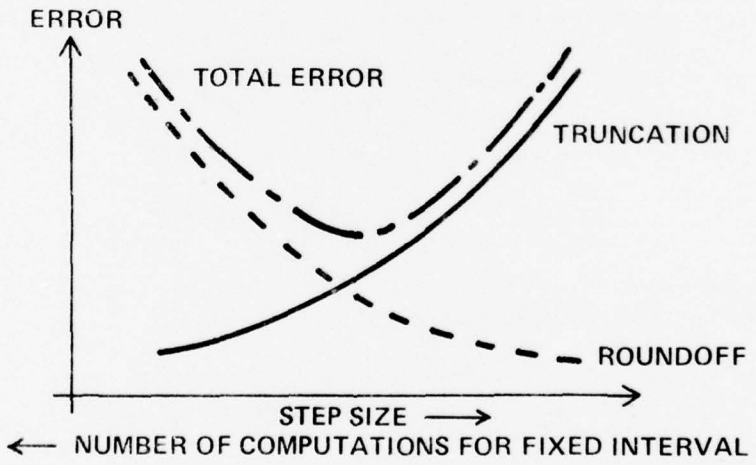


Figure 2. Error vs Step Size

2.2 An Important Aside

We have just derived the formula for Simpson's Rule by truncating the Newton interpolation formula. Unless the function $\dot{y}$ itself goes to infinity over the interval $[t_0, t_2]$, the error in the approximation will be finite,

and the error over a finite number of intervals will also be finite. The integration of $\dot{y}$, however, is here a quadrature; a given $\dot{y}$, explicitly defined as a function of t, has been approximated over a known interval. In the more general problem with dynamics, $\dot{y}$ will depend on y, and errors in the computation of y will be fed back to propagate and perhaps grow without bound. In such a situation, an unbounded error can arise in a finite interval.

The question of such growth in errors is a question of stability, and we will see later how the control system perspective provides a very convenient way to examine the stability of a numerical integrator. Note for now that many integrators, of which Simpson's is one, may be acceptable for quadrature problems but are patently unstable for the dynamic problem.

2.3 Errors From the Control System Point of View

From the discussion above, we can see how the classical approach leads us to the use of the integration step size or the order of the integrator (i.e., the degree of truncation) to the control of errors. Means are available to evaluate bounded magnitudes for the errors, and can be found in many references [1],[10],[11]. When one uses variable-step-size or variable-order algorithms to control those bounds, one pays the penalty of problems in increased complexity and increased computational burden as the order increases. Tunable integration and the control-theory analysis provide an alternative approach to error control, one in which neither step size nor order are varied. The parameters that tune the integrator control the error. Avoided are the restart problems of changing step size (for the multistep methods) and the complexity of higher order.

To discuss errors from the control-system perspective, we must look at the integrator itself as a control system. If we define a control system as a collection of components and/or algorithms designed to take a given, known or uncertain, input signal and from it produce an output with certain desired properties, then the function of an integrator is clear: the numerical integrator is a linear-transfer-function control system that takes an input integrand and produces as its output the integral of the input. Schematically we can portray this function as in Figure 3.

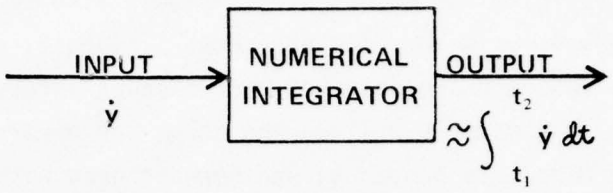


Figure 3. The Numerical Integrator as a Control System

The true integrator would produce the true integral at the output. How far from the true integrator the numerical integrator is can be analyzed in terms of the transfer function (using the Laplace variable s):

$$G(s) = \frac{\text{output}(s)}{\text{input}(s)}$$

There are two fundamental properties of G(s), the analyses of which comprise the heart of this report. These properties are 1) the locations of its poles and zeros, and 2) the magnitude and phase characteristics of its frequency response. How these properties differ from those of the true integrator is a direct measure of the performance of our numerical integrator.

2.4 The True Integrator

In the Laplace domain the transfer function of the true integrator is

$$G_t(s) = \frac{1}{s} \qquad (2\text{-}5)$$

This single pole at the origin maps into a single pole at (1,0) in the z-plane, where $z = e^{sT}$. Thus, pole-zero maps of the true integrator in the s- and z-planes are simply as shown in Figure 4.

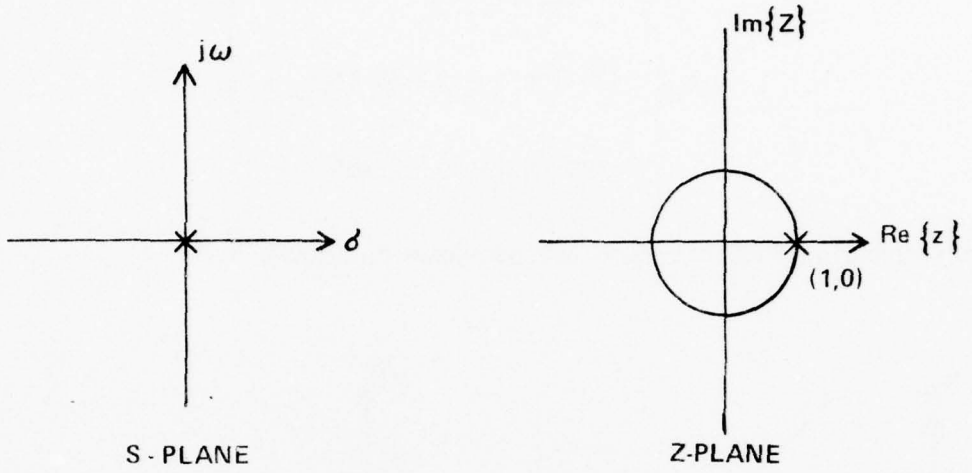


Figure 4. True Integrator Pole Location†

We should like any numerical algorithm to have a pole close to that of the true integrator. We should also like as few additional poles in the algorithm as possible, for additional poles result in additional lag which tends to be destabilizing, and they add to the complexity of the algorithm.

A second basic property of the true integrator is its frequency response magnitude and phase. Substituting $j\omega$ for s in Eq (2-5), we have

$$G_t(j\omega) = \frac{1}{j\omega} = -\frac{j}{\omega} \qquad (2\text{-}6)$$

The magnitude and phase are

---

†It is interesting to note that simply taking the z-transform of $\frac{1}{s}$ would give $G_t(z) = Z\left\{\frac{1}{s}\right\} = \frac{z}{z-1}$. The corresponding difference equation is $x_{n+1} = x_n + \dot{x}_{n+1}$, which implies a unit integration step in order to have any practical meaning.

$$\left.\begin{aligned} M &= |G_t(j\omega)| = \frac{1}{\omega} \\ M_{db} &= 20 \log_{10} M = -20 \log_{10} \omega \end{aligned}\right\} \qquad (2\text{-}7)$$

$$\phi = \text{Arg}\ \{G_t(j\omega)\} = -90^{o} \qquad (2\text{-}8)$$

The corresponding Bode diagrams are as shown in Figure 5.

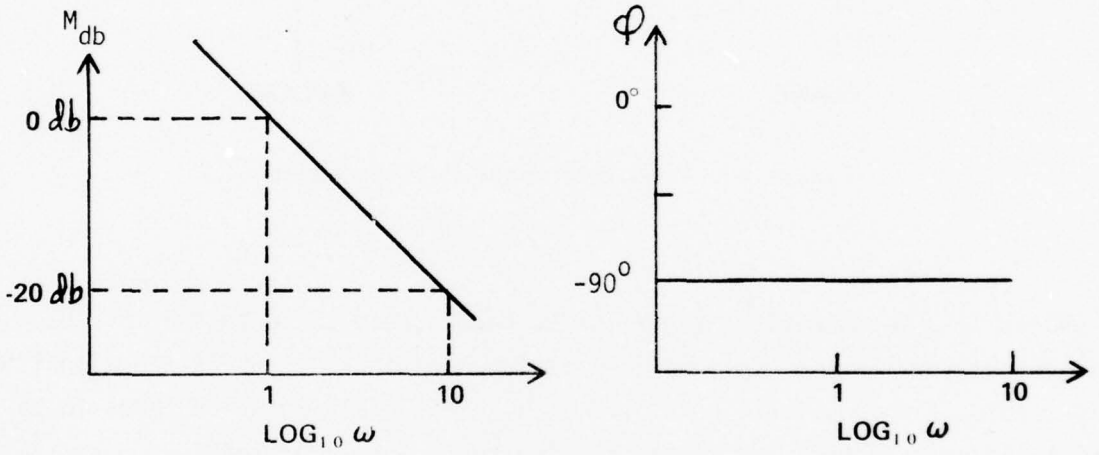


Figure 5. True Integrator Frequency Response

Any numerical algorithm should have magnitude and phase characteristics closely approximating the 20 db/decade gain slope and 90^{o} of phase lag, at least over the frequency range of interest.

These two fundamental properties of pole location and frequency response provide a control system perspective for the analysis of numerical integration techniques that will be used throughout the remainder of this report. Variations in pole location, additional poles and zeros, variation of M from $1/\omega$, and variation of ϕ from -90^{o} are all direct measures of the error in the numerical integration algorithm.† We will later see a clear tie between

†The use of these criteria is not unfamiliar to the engineer involved in simulation.[12],[13]

phase and gain distortion and the classical concept of truncation error. This will arise from the use of holding circuits to reconstruct sampled signals. The reconstructed signals are simply polynomial approximations to their continuous predecessors.

2.5 The "Ideal" Integrator†

Having defined what will be our control system perspective, we can derive a set of ideal integrators that have the exact properties desired, but only for the design input. These ideal integrators take specified inputs and transform them into the corresponding known integrals for their output.†† For example, consider an ideal cosine integrator. The z-transforms of input and output are

$$Z\{\cos \omega_d t\} = \frac{z\,(z - \cos \omega_d T)}{z^2 - (2 \cos \omega_d T)\, z + 1}$$

$$Z\{\sin \omega_d t\} = \frac{(\sin \omega_d T)\, z}{z^2 - (2 \cos \omega_d T)\, z + 1}$$

where ω_d is the design frequency. The ideal transfer function is

$$G_i(z) = \frac{1}{\omega_d} \frac{Z\{\sin \omega_d t\}}{Z\{\cos \omega_d t\}} = \frac{\sin \omega_d t}{\omega_d\,(z - \cos \omega_d T)} \qquad (2\text{-}9)$$

The Bode magnitude and phase are found by substituting $e^{j\omega T}$ for z in Eq (2-9):

$$G^*(j\omega) = \frac{\sin \omega_d T}{\omega_d\,(e^{+j\omega T} - \cos \omega_d T)} \quad \text{†††}$$

---

†This section is based on ideas proposed and initially developed by Ronald E. Janosko.

††This concept is not too dissimilar from the ideas presented by Fowler.[14]

†††Note that we have the sampled frequency response $G^*(j\omega)$ and not the continuous-system response $G(j\omega)$.

$$\left.\begin{aligned} M &= \frac{\sin \omega_d T}{\omega_d} \left[1 + \cos^2 \omega_d T - 2 \cos \omega T \cos \omega_d T\right]^{-\frac{1}{2}} \\ \phi &= -\tan^{-1} \left\{ \frac{\sin \omega T}{\cos \omega T - \cos \omega_d T} \right\} \end{aligned}\right\} \qquad (2\text{-}10)$$

At the design frequency ω_d, these relations reduce to

$$M = \frac{1}{\omega_d}$$

$$\phi = -90^{\circ}$$

Thus, at the design frequency we have the desired frequency-response characteristics, independent of the time step selected. There is also only one pole, though its location will depend upon the value of $\omega_d T$.

What happens, however, if we use this ideal integrator at other than the design frequency? If we let $\omega_d T = \pi/6$ (the sampling ratio is thus 1/12; i.e., $\omega_d/\omega_s = \omega_d/(2\pi/T) = 1/12$), Eqs (2-9) yield the data shown in Table 1.

Table 1: Ideal Cosine Integrator Frequency Response

ω/ω_d	$\omega_d \cdot M$	$\omega \cdot M$	ϕ(deg)
0.5	1.80	0.90	- 69
0.7	1.37	0.96	- 79
1.0	1.0	1.0	- 90
1.3	0.79	1.02	- 98
1.5	0.69	1.04	-103

Both the magnitude and phase characteristics degrade as we move away from ω_d. Note that both products, $\omega_d \cdot M$ and $\omega \cdot M$, drift away from unity, which the true integrator produces at all frequencies.

We noted above that the frequency response at the design frequency was independent of T. Concomitantly, the accuracy is independent of T: the

ideal integrator, not surprisingly, gives exact results regardless of step size. This can be seen by writing the difference equation for the integrator. From Eq (2-9) that difference equation is

$$x_{n+1} = ax_n + b\dot{x}_n$$

where

$$a = \cos \omega_d T$$

$$b = \frac{1}{\omega_d} \sin \omega_d T$$

Given the initial conditions $\dot{x}_n = \cos n\omega_d T$ and $x_n = \frac{1}{\omega_d} \sin n\omega_d T$, we algebraically obtain the result $x_{n+1} = \frac{1}{\omega_d} \sin [(n+1)\omega_d T]$, independent of T. Using the same example as was used to generate Table 1, we can perform the numerical integration and see the actual errors. Let $\omega_d = 1$ and $T = \pi/6$ so that $\omega_d T = \pi/6$ as before. The data in Table 2 show ten integration steps each for $\omega = 0.5$, $\omega = 1.0 = \omega_d$, and $\omega = 1.5$.

Table 2: Ideal Cosine Integrator Performance

t	$\omega = 0.5$		$\omega = 1.0$		$\omega = 1.5$	
	x_t[1]	x_c[2]	x_t[1]	x_c[2]	x_t[1]	x_c[2]
0	0	0	0	0	0	0
0.524	0.518	0.500	0.500	0.555	0.471	0.500
1.047	1.000	0.916	0.866	0.866	0.667	0.787
1.571	1.414	1.226	1.000	1.000	0.471	0.681
2.094	1.732	1.416	0.866	0.866	0.000	0.236
2.618	1.932	1.476	0.500	0.500	-0.471	-0.295
3.142	2.000	1.408	0.000	0.000	-0.667	-0.609
3.665	1.932	1.219	-0.500	-0.500	-0.471	-0.528
4.189	1.732	0.926	-0.866	-0.866	0.000	-0.103
4.712	1.414	0.552	-1.000	-1.000	0.471	4.410
5.236	1.000	0.125	-0.866	-0.866	0.667	0.709

Notes: 1. $x_t(t_n) = \frac{1}{\omega} \sin \omega t_n$: true integrator result

2. $x_c(t_{n+1}) = ax_c(t_n) + b\dot{x}_c(t_n)$: ideal cosine integrator result

$\dot{x}_c(t_n) = \cos \omega t_n$; $a = \cos \omega_d T$; $b = \frac{\sin \omega_d T}{\omega_d}$

The relation between the data in the two tables is clearly seen in the case of $\omega = 0.5$. The magnitude is too small and the phase lag is too little in the first table. The peak seen in Table 2 is less than the true peak and occurs before it should, just as the frequency response data indicated. For $\omega = 1.5$ the larger-than-ideal magnitude is evident, but the excessive phase lag cannot be seen due to the spacing of the data points.

Some additional ideal integrators are shown in Table 3.

Table 3: Ideal Integrators

Input	G(z)
$u(t)$	$T/(z-1)$
t	$T/2 \quad \frac{z+1}{z-1}^{*}$
e^{at}	$\frac{1/a\ (e^{aT} - 1)}{z-1}$
$\sin \omega_d t$	$\frac{1 - \cos \omega_d T}{\omega_d \sin \omega_d T} \cdot \frac{z+1}{z-1}$
$\cos \omega_d t$	$\frac{\sin \omega_d T}{\omega_d} \cdot \frac{1}{z - \cos \omega_d T}$

*Tustin's Transformation

From this elementary analysis, admittedly using an example of quadrature to make our points, it is clear how one can apply the methods of the control engineer to the analysis of a numerical integrator. We have seen how the ideal integrator may be derived from the perspective of a linear transfer function. Its range of usefulness is, however, very limited. If we define a sampling ratio r as follows

$$r = \omega/\omega_s, \quad \omega_s = 2\pi/T \tag{2-11}$$

then, for the example given earlier, $\omega = 0.5$ corresponded to $r = 0.04$, $\omega = 1$ to $r = 0.08$, and $\omega = 1.5$ to $r = 0.13$. We shall see in Chapter 3 that this is indeed a very limited range of the sampling ratio. If one were to design an integrator from the ground up using a control-theory perspective, a prime consideration would be making the useful range of r as large as possible. This is precisely what has been done in the development of the tunable integrator which we will discuss in depth beginning with the next chapter.

CHAPTER 3
Introduction to Tunable Integration

This chapter begins the heart of the report: the analysis of tunable integration (TI). In it we will briefly review the formulation of TI, which is the brainchild of Jon M. Smith.[4],[5] We will derive the zero-order-hold TI (ZOH-TI) formula and will demonstrate its application to a second-order differential equation. The following two chapters will present frequency-response and root-locus analyses, respectively.

3.1 Basic Formulation[5]

In Smith's words, the approach to developing the TI is one of "Synthesizing a discrete approximation to continuous integration. . ."[5] To do this, four basic components are required: 1) samplers to represent the discretization of the entire process of digital computation; 2) a reconstructor to provide a continuous signal from the sampled input; 3) a compensator to enable control of the distortion introduced by both the sampling and reconstruction processes; and 4) an integrator to integrate the compensated signal. Given these four components, there are two ways to construct the discrete integrator. The first is to use continuous compensation as shown in Figure 6a, and the second is to use discrete compensation as shown in Figure 6b.

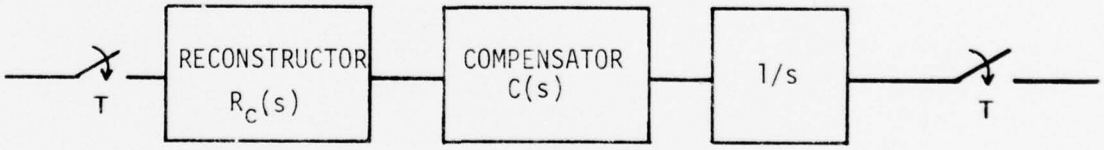


Figure 6a. Continuous Compensation

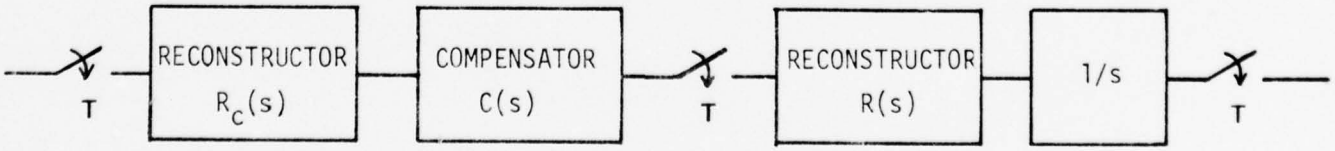


Figure 6b. Discrete Compensation

Two families of integrators have been generated by Smith depending upon the type of compensation. In general, for the same compensator, the continuous-compensation integrators are less complex. Formulas for both types are presented in [5]. Reference [6] develops a set of continuous-compensation formulas that are less complicated than Smith's. In this report I will discuss almost exclusively the ZOH-TI. Let us now look in detail at the reconstructor and compensator.

3.1.1 The Reconstructor

As stated, the reconstructor provides a continuous signal from the samples it is provided. The complexity of the reconstructed signal, and its amplitude and phase distortion, will be a function of the order of the reconstructor. That order is essentially the order of the polynomial that is fit to the sample points. Since we are here fitting polynomials to data points, this process is related to the process typically followed in classical developments of numerical integrators. The difference arises from our control-system perspective and the use of a compensator which allows us to control the errors induced by the polynomial approximation.

The zero-order-hold reconstructor provides a stair-step approximation to a continuous signal as shown in Figure 7.

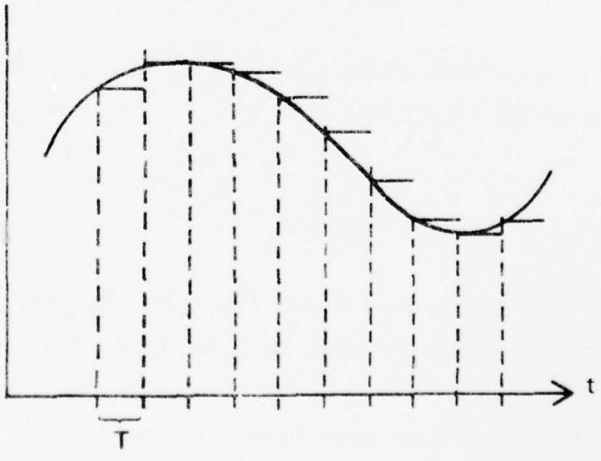


Figure 7. Stair-Step Approximation

If we examine the response to a single sample, we see that it is simply a rectangular pulse of magnitude equal to the sample value and of duration T. Since the sample is ideally an impulse, that rectangle is the impulse response of the zero-order hold as seen in Figure 8 (for a unit impulse).

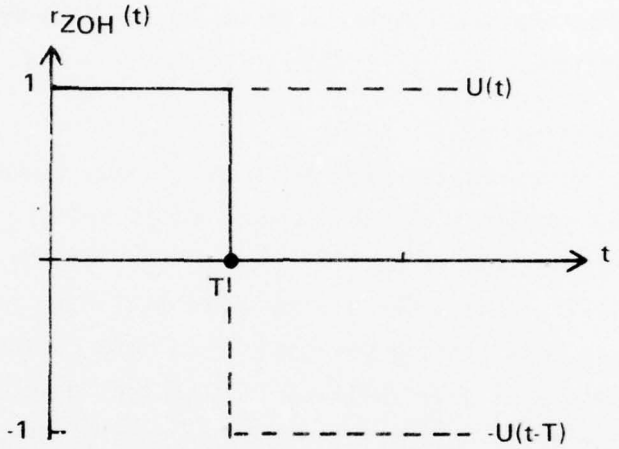


Figure 8. Zero-Order Hold Impulse Response

The figure also shows that the impulse response $r_{ZOH}(t)$ is made up of a positive step function at $t = 0$ and a negative step at $t = T$:

$$r_{ZOH}(t) = u(t) - u(t-T) \tag{3-1}$$

Taking the Laplace transform of Eq (3-1) gives the standard expression for a zero-order-hold transfer function:

$$R_{ZOH}(s) = \frac{1-e^{-sT}}{s} \tag{3-2}$$

We could take the z-transform of R and find that $R_{ZOH}(z) = 1$, but that helps little since the transformation that must be taken for continuous-compensation TI is $Z\{R_c C \frac{1}{s}\}$.

Other reconstructors that have been used to develop TI formulas are first- and second-order and triangular holds, the last not being physically realizeable but still being mathematically useful. While capable of improving

accuracy, high-order holds add complexity and additional poles, both being contrary to the objectives of simplicity and minimization of extraneous poles. This is one of the primary reasons that this report deals primarily with the ZOH-TI.

3.1.2 The Compensator

The function of the compensator is to enable control of the distortion introduced by the sampling and reconstruction processes. Rather than utilize the standard form of lead-lag compensators, Smith employs the complex exponential form

$$C(s) = \lambda e^{\gamma sT} \tag{3-3}$$

The two parameters λ and γ will be seen later to be the gain and phase compensation parameters of the TI. If we let $s = \sigma + j\omega$, then

$$|C(j\omega)| = \lambda e^{\gamma\sigma T}$$

$$\text{Arg } \{C(j\omega)\} = \gamma\omega T$$

These relations show us that while γ primarily affects phase compensation, it also affects the gain. We will see this explicitly in the difference equation for the ZOH-TI.

In employing the compensator, Smith absorbs $e^{\gamma sT}$ into the product $R_c C \frac{1}{s}$, transforms back to the time domain and then transforms the result to the Z-domain. I have, instead, expanded the exponential and truncated after the second term (admittedly creating more truncation error, but the control perspective allows me to monitor and counter the deliterious effects). It is in this manner that the continuous-compensation integrators of [6] were developed. The compensator used for the remainder of this report is thus

$$C(s) = \lambda(1 + \gamma sT) \tag{3-4}$$

It is interesting that both Smith's approach and my approach give the same result for the ZOH-TI. The discrete compensator used by Smith is developed using a triangular hold and the compensator of Eq (3-4). When used as in [5]

with R_i (see Figure 6b) being a zero-order hold, the result is again the same form as the continuously compensated ZOH-TI.

3.2 Development of the Zero-Order-Hold Tunable Integrator

3.2.1 The Transfer Function

Using continuous compensation, a zero-order-hold reconstructor and the truncated polynomial compensator, we can redraw Figure 6a as follows:

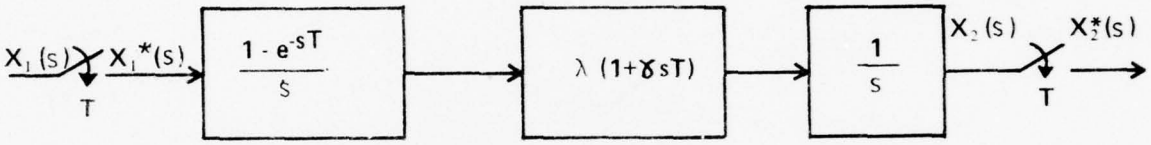


Figure 9. ZOH-TI

Note that the input $X_1(s)$ is sampled to provide $X_1^*(s)$, and the continuous output $X_2(s)$ is sampled to provide $X_2^*(s)$. The transfer function of the continuous path between the samplers is simply the product of the three elements in the path:

$$\frac{X_2(s)}{X_1^*(s)} = \frac{1-e^{-sT}}{s} \lambda(1+\gamma sT) \frac{1}{s}$$

By taking the z transformation of this, we obtain the total transfer function between the input and output of this discrete integrator. Initially we can write

$$G(z) = Z\left\{\frac{X_2}{X_1}\right\} = \lambda \frac{z-1}{z} Z\left\{\frac{1+\gamma sT}{s^2}\right\}^{\dagger}$$

From a set of transform tables such as those in [10] and [11], we have the result

---

†Though no sampler exists between the reconstructor and compensator, the mathematics of z-transforms allows us to factor out the term $(1-e^{-sT})$, the transformation of which is $(z-1)/z$.

$$\left.\begin{aligned} G(z) &= \frac{\lambda T[\gamma z + (1-\gamma)]}{z-1} \\ &= \frac{p(z-q)}{z-1} \end{aligned}\right\} \qquad (3\text{-}5)$$

where

$$\left.\begin{aligned} p &= \lambda T \gamma \\ q &= \frac{\gamma - 1}{\gamma} \end{aligned}\right\} \qquad (3\text{-}6)$$

Equation (3-5) is the transfer function of the ZOH-TI. Equations (3-6) reiterate the earlier point that γ affects both the phase and gain of the integrator. More will be said about the frequency response and root-locus characteristics of this integrator in the next two chapters, but for the moment we can note that the pole-zero map of the basic integrator (an open-loop device) is as shown in Figure 10.

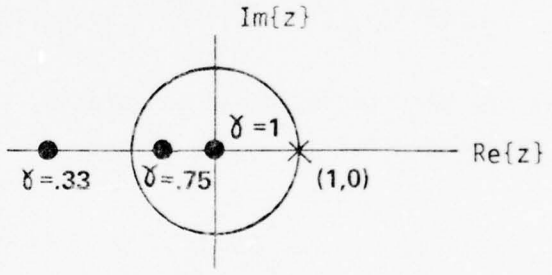


Figure 10. ZOH-TI Pole-Zero Map

There is a pole at (1,0) matching the true integrator shown in Figure 4. There is also a zero whose location depends on γ according to the relation for q in Eq (3-6). For $\gamma \geq \frac{1}{2}$ the zero is always within or on the unit circle, while for $\gamma < \frac{1}{2}$ the zero will always be outside the unit circle (to the left for $\gamma > 0$ or to the right for $\gamma < 0$, though the desirability of ever using $\gamma < 0$ is very doubtful due to the excessive lag that would be induced). If a loop were placed around the integrator to form a simple, first-order system, then increasing the forward path gain, p, by increasing T would drive the pole to the zero. For $\gamma \geq \frac{1}{2}$ the pole would never exit the unit circle and the

system would never go unstable (it would be inaccurate, however), while for $\gamma > \frac{1}{2}$, sufficiently large integration steps would result in instability. This will be clearly evidenced in plots to appear later in this chapter.

This brief analyses demonstrates the straight-forward manner in which a numerical integrator can be studied using control system methodology. The question of stability will be examined in more detail later, and the relation between an examination of the actual integrator pole locations and the more familiar, to numerical analysts, examination of the equation for propagation of the integrator error will be shown.

3.2.2 The Difference Equation

From G(z) we can readily find the difference equation that represents the algorithm to be programmed on a digital computer. Eq (3-5) yields the relation

$$X_2(z) = G(z)X_1(z)$$

$$(z-1)X_2(z) = \lambda T[\gamma z + (1-\gamma)]X_1(z) \tag{3-7}$$

Since our objective has been to design an integrator, it is natural that x_1 be representative of $\dot{x}$ and x_2 of x. Hence we say

$$\frac{x_2}{x_1} \approx \frac{x}{\dot{x}}$$

and rewrite (3-7) as

$$(z-1)X(z) \approx \lambda T[\gamma z + (1-\gamma)]\dot{X}(z) \tag{3-8}$$

We will, in fact, designate equality when we program Eq (3-8). The measure of our efforts will be how well x does represent the true integral of $\dot{x}$. Noting that $zX_n = X_{n+1}$, we write the desired difference equation of the ZOH-TI from Eq (3-8) in the following form:

$$\boxed{x_{n+1} = x_n + \lambda T[\gamma \dot{x}_{n+1} + (1-\gamma)\dot{x}_n]} \tag{3-9}$$

The ZOH-TI is an implicit integrator, requiring $\dot{x}_{n+1}$ to evaluate x_{n+1}. Therefore, a predictor-corrector procedure or an extrapolation of $\dot{x}_n$ is required. The effect of γ is to weight the leading derivative $\dot{x}_{n+1}$ versus the lagging derivative $\dot{x}_n$. Recalling our previous discussion of the effect of γ on the location of the integrator zero, we can intuitively say that $\gamma < \frac{1}{2}$ tends to be a lag situation (destabilizing) and $\gamma > \frac{1}{2}$ tends to be a lead situation (stabilizing). This lead-lag effect of γ will become very clear when we look at the frequency response in Chapter 4.

3.2.3 Phase-Shifted Classical Integrators

Smith has observed that "many of the widely varied classical numerical integration formulas . . . are actually the same integrator, differing only in the amount of phase shift of the integrand."[5] Letting γ take on the values 0, $\frac{1}{2}$, 1 and 3/2, we obtain four familiar integration formulas from Eq (3-9) as shown in Table 4.

Table 4: The ZOH-TI and Classical Integrators

γ	Difference Equation ($\lambda = 1$)	Name
0	$x_{n+1} = x_n + T\dot{x}_n$	Euler
$\frac{1}{2}$	$x_{n+1} = x_n + \frac{T}{2}(\dot{x}_{n+1} + \dot{x}_n)$	Trapazoidal
1	$x_{n+1} = x_n + T\dot{x}_{n+1}$	Rectangular
3/2	$x_{n+1} = x_n + \frac{T}{2}(3\dot{x}_{n+1} - \dot{x}_n)$	Adams 2nd Order

3.3 An Example Application[7]+

Consider a damped second-order oscillator represented by the differential equation

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = \omega_n^2 f(t) \qquad (3\text{-}10)$$

+This is the same example as examined by Smith in [4] and [5]. In [7], I explain the reasons for differences in our results.

where ω_n is the natural frequency, ζ is the damping coefficient and f(t) is a forcing function. The Laplace transform of this equation is

$$s^2X(s) + 2\zeta\omega_n sX(s) + \omega_n{}^2X(s) = \omega_n{}^2F(s)$$

We can draw the block diagram of this system as in Figure 11.

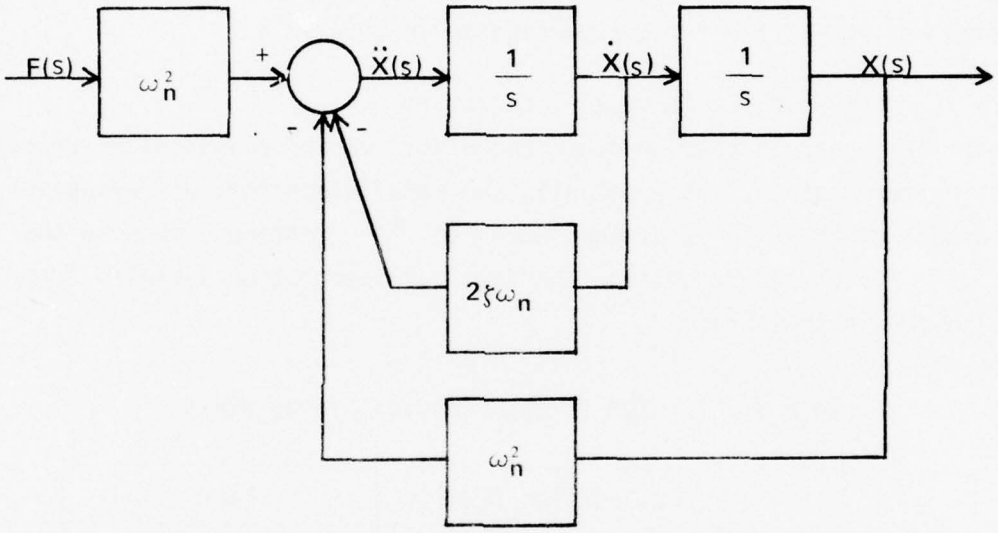


Figure 11. Second-Order System Block Diagram

The ZOH-TI was implemented in the following manner:

1. $\dot{x}_{n+1}$ and x_{n+1} were initially assumed equal to $\dot{x}_n$ and x_n, respectively. This is equivalent to using an Euler predictor.

2. Eq (3-9) was used to compute x_{n+1}, which was, in turn, used to evaluate $\dot{x}_{n+1}$. The process was iterated until convergence.

An analysis of block diagrams will be done in Chapters 5 and 6, but at this point it should be noted that the final block diagram of the software system is not obtained by simply replacing each integrator in Figure 11 with a ZOH-TI.

Tests have been run for a number of forcing functions (including step, impulse, and sinusoidal) and the results are similar for each. We will here present only the case of sinusoidal forcing at resonance; i.e., $f(t) = \sin \omega_n t$.

For this forcing function, the true solution is found by analytic means to be

$$x(t) = \tfrac{1}{2}\, e^{-\zeta\omega_n t}\left[\frac{1}{\zeta}\cos\omega_d t + \frac{\omega_n}{\omega_d}\sin\omega_d t\right] - \frac{1}{2\zeta}\cos\omega_n t \qquad (3\text{-}11)$$

where $\omega_d = \omega_n\sqrt{1-\zeta^2}$ is the damped frequency. The plots that follow were computed using $\omega_n = 1$ Hz, $\zeta = 0.3$, and $\lambda = 1$. The integration step T and phase-tuning parameter γ were varied.

From Eq (3-11) we note that the true solution is sinusoidal, having an expanding, exponential envelope with a steady state magnitude of $\frac{1}{2\zeta} = 1.67$. Figures 12, 13 and 14 portray the results of solving the differential equation (3-10) with the ZOH-TI tuned to three values of γ: 0, 1, and ½, respectively.

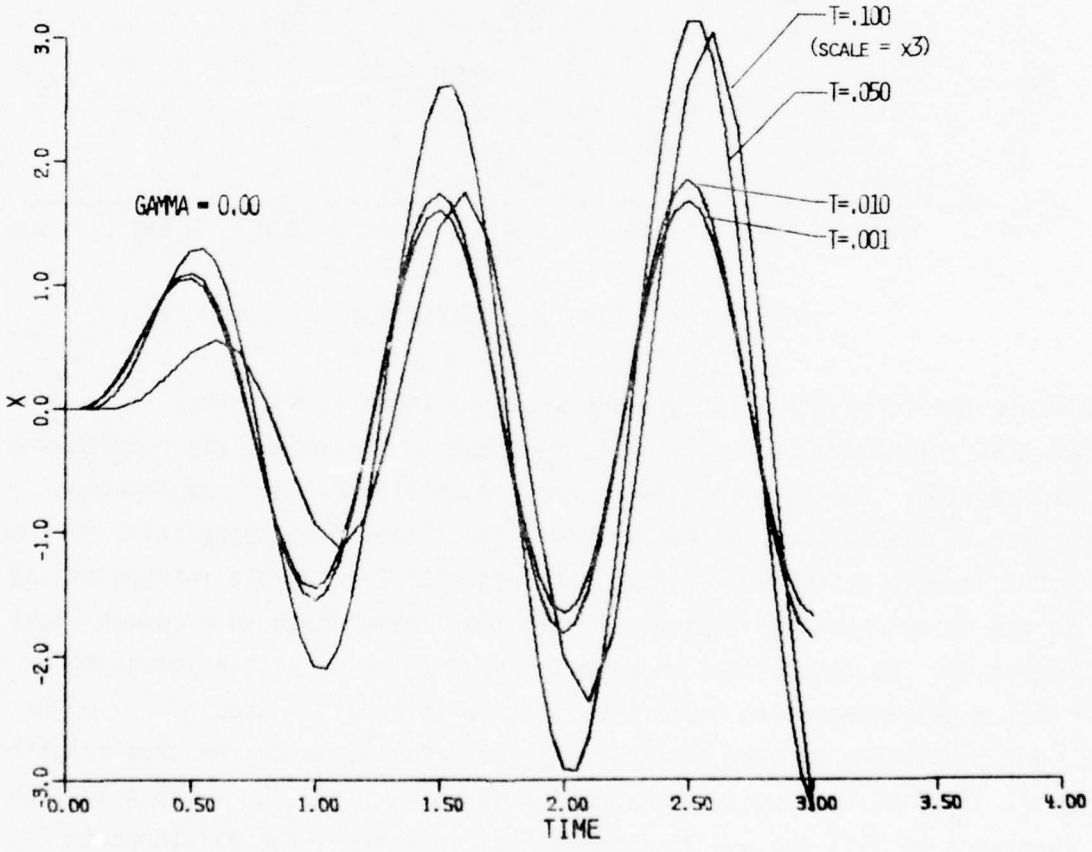


Figure 12. ZOH-TI Solution for $\gamma = 0.0$

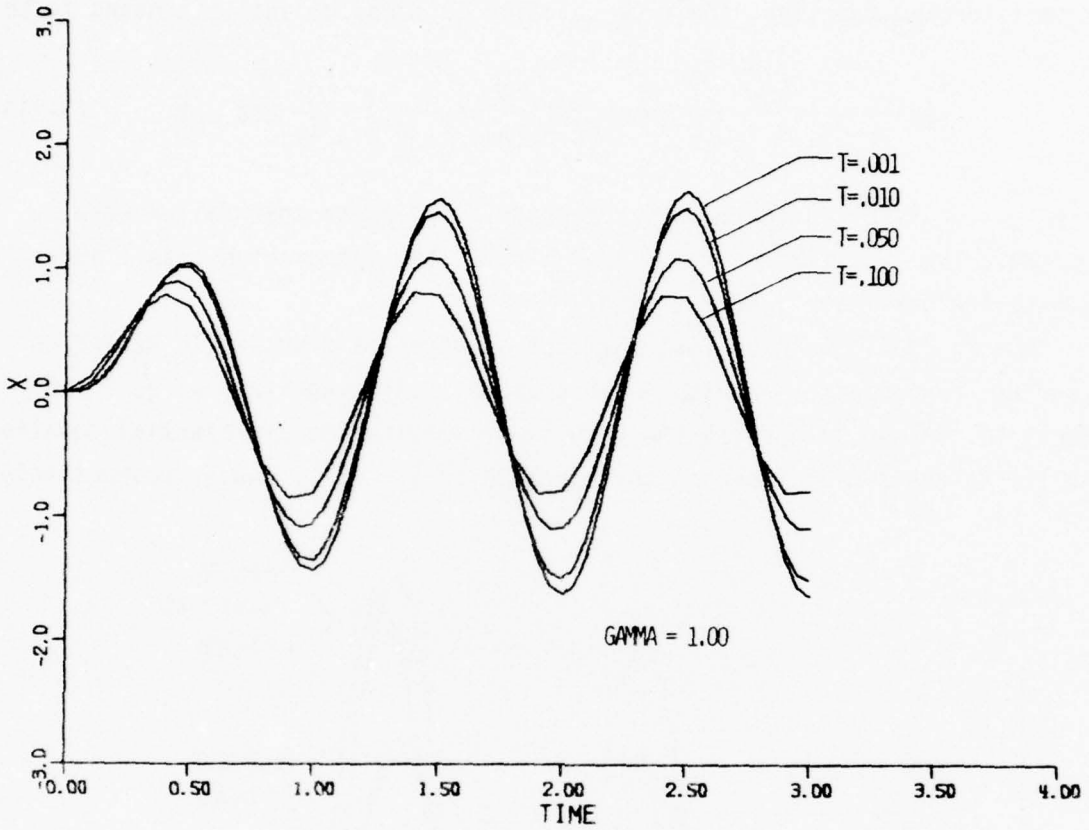


Figure 13. ZOH-TI Solution for $\gamma = 1.0$

These are the first three integrators shown in Table 4, and since it is well known that trapazoidal integration is the best of the three, the results are not surprising. The significance of these data lies in their explanation according to the laws of control theory. It is from that perspective that we gain the insight as to what we have obtained with the tunable integrator and as to how we can view all methods of numerical integration in a common light. So forget for the moment that we know these as three classical integrators.

All three integrators perform well at an integration step of T = 0.001 sec, and the curves plotted at that step size closely match the true solution of Eq (3-11). At that step, the sampling ratio is $r = \omega_d/\omega_s = .95 \times 10^{-3}$ and in Chapter 4 we will see why it is perfectly reasonable for all three to produce similar results at that step size. In Figure 12, $\gamma = 0$, and we have

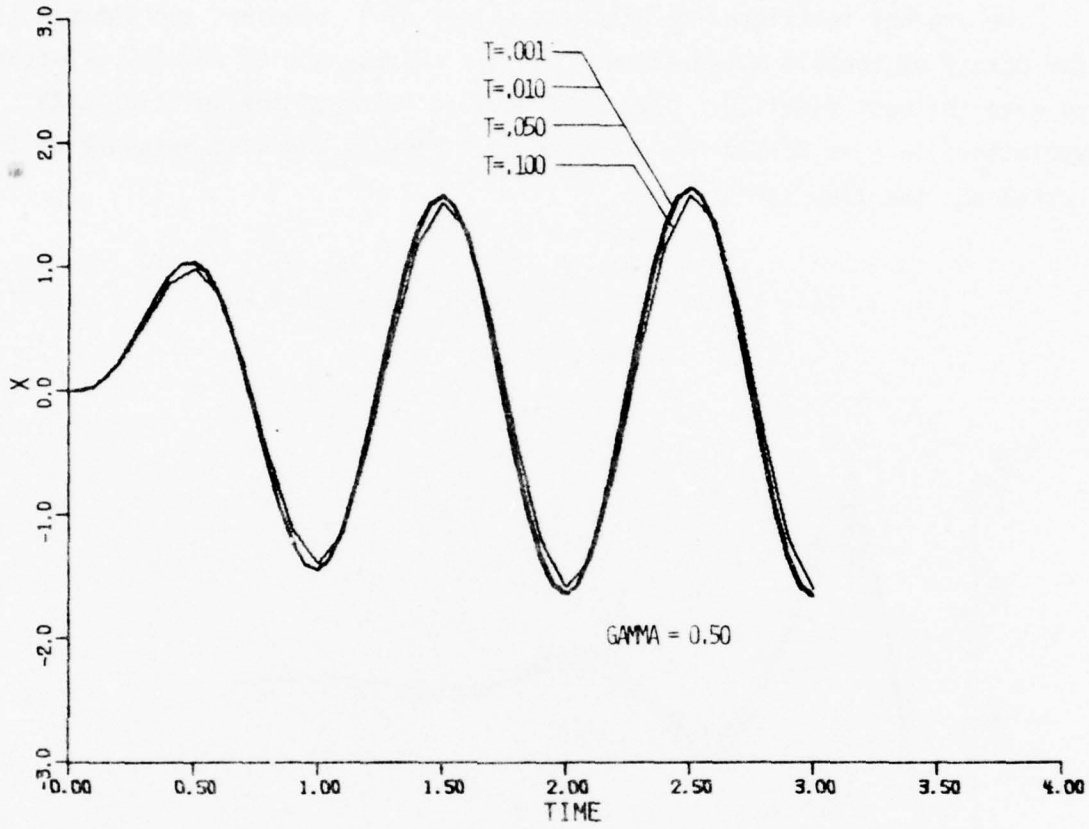


Figure 14. ZOH-TI Solution for $\gamma = 0.5$

intuitively noted that this situation tends to produce a lot of phase lag. In fact, as we increase T the amount of lag increases and the peaks are delayed and the solution diverges, having completely blown up at T = 0.100. (Note the change in scale for that curve.) Quite the opposite occurs in Figure 13 where $\gamma = 1.0$. Our intuitive feeling of lead is buttressed by the progressively earlier occurrence of the peaks as T increases. Rather than going unstable, however, sensitivity is lost and the solution decreases in magnitude, producing a bounded rather than unbounded error. Note that the shift of γ from zero to unity has completely changed the sensitivity of the integrator to increases in step size. The third figure of the sequence shows the results for a balance between lead and lag. When $\gamma = 0.50$ we can increase the integration step by two orders of magnitude and still have a faithful representation of the true solution.

We are not restricted to discrete values of γ, however, and therein lies the beauty of tunable integration. In this example, a γ of about 0.52 appears to give the best fidelity. The continuous shifting of the solution with variations in γ is displayed in Figure 15. This is the step response of the system and the true solution is

$$x(t) = u(t) - e^{-\zeta\omega_n t}\left(\cos \omega_d t + \zeta \frac{\omega_n}{\omega_d} \sin \omega_d t\right) \qquad (3\text{-}12)$$

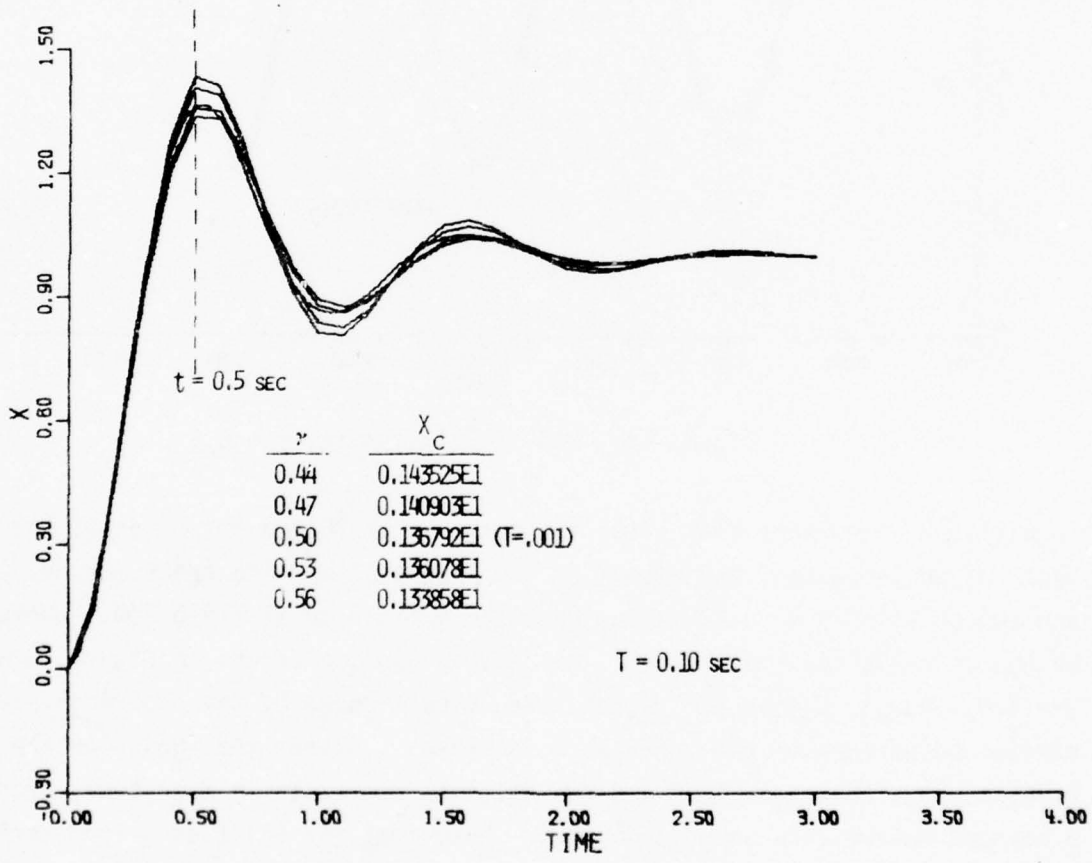


Figure 15. ZOH-TI Step-Response Solution

The true solution at t = 0.5 sec is x(0.5) = 1.367918 and this is seen to be accurately reproduced by the curve computed with γ = 0.5 and T = 0.001. The

remaining curves are computed for the values of γ indicated on the figure and T = 0.100. For γ= 0.53 the solution matches the true value to three decimal places at a time step two orders of magnitude larger than .001.

3.4 The Advantages of Tunable Integration

In this chapter we have synthesized a discrete approximation to the continuous process of integration. This discrete numerical integrator possesses the controls necessary to manipulate its phase and amplitude characteristics in a manner that allows tuning the system for the best results. The ability to tune the integrator increases its region of stability and accuracy at larger time steps (see the discussion in Chapter 5 on stability) thereby allowing for increased computational speed and decreased computational cost.

CHAPTER 4

Frequency-Response Characteristics of the ZOH-TI

4.1 Why the Frequency Response

In Chapter 2 we saw how the performance of a numerical integrator could be assessed by comparison of its magnitude and phase-angle characteristics with those of the true integrator as expressed by Eqs (2-7) and (2-8) and portrayed in Figure 5. For this comparison we need to evaluate the frequency response of the numerical integrator. This has been done in [6] for several continuously and discretely compensated tunable integrators. The analysis of only the ZOH-TI will be presented here. Rosko[12] looks at the frequency characteristics of a number of "discrete integrators," such as Tustin and Boxer-Thaler, and Tou[13] briefly looks at a number of quadrature formulas, such as rectangular and Simpson.

There is a point worth noting before we proceed, and it is implied by Tou's use of the word quadrature. In the three referenced analyses, the frequency response of the open-loop integration formula is evaluated. The information so garnered cannot be used indiscriminantly in the evaluation of software for the solution of differential equations, for, in determining the solution, the dynamics of the differential equation and the manner of implementing the algorithm are important. This open-loop frequency-response data must be used with other information to get a full evaluation of the integrator, but it is a useful and important part of the overall picture of integrator performance.

4.2 Methods of Evaluation

There are a number of approaches possible to obtain the desired data, and the choice of which to use is relatively free. Three analytic methods available are as follows. One can simply make the substitution $z = e^{j\omega T}$, as do Rosko and Tou, and obtain $G^*(j\omega)$ directly from $G(z)$. Smith[5] takes an alternative approach to the same end. He makes the transformation $z = \frac{1+w}{1-w}$ where $w = j \tan(\omega T/2)$. (A little algebra shows that this relation for z is equivalent to $e^{j\omega T}$.) The third analytic approach involves obtaining the actual output in the z-domain by multiplication of $G(z)$ and the z-transform of the

input and then taking the inverse transformation of the result. Determination of the magnitude and phase is then by inspection; e.g.,

$$A \sin(\omega t + \phi) = Z^{-1}\left\{Z\{\sin \omega t\}G(z)\right\} \tag{4-1}$$

If one disdains algebra, the frequency response can be determined by actually integrating a sinusoidal signal and evaluating the amplitude and phase of the resultant output. This numerical approach is fraught with a number of problems, principally with respect to interpolation between data points to find peaks and cross-overs as the sampling ratio increases. There is also a bias generated in the solution, which must be subtracted out before any evaluation is made. The third analytic method is very tedious, but it does enable an analytic attack on evaluating the bias in the numerical computation. The remaining two approaches are equivalent and either should be preferable to the novice. The method used by Tou and Rosko will be employed in what follows.

4.3 Frequency Response of the ZOH-TI

To find $G^*(j\omega)$, we begin by writing the second of Eqs (3-5) in the following form:

$$G(z) = \frac{p(1-qz^{-1})}{1-z^{-1}}$$

Now make the substitution $z^{-1} = e^{-j\omega T}$ and obtain

$$G^*(j\omega) = p\,\frac{(e^{+j\omega T/2} - qe^{-j\omega T/2})}{e^{j\omega T/2} - e^{-j\omega T/2}} \tag{4-2}$$

From Euler's equation, we know that

$$\sin \omega T/2 = \frac{e^{j\omega T/2} - e^{-j\omega T/2}}{2j}$$

Hence, we can write Eq (4-2) as

$$G^*(j\omega) = (-jp/2)\,\frac{(e^{j\omega T/2} - qe^{-j\omega T/2})}{\sin \omega T/2}$$

Expanding the remaining exponentials, we obtain the result

$$G^*(j\omega) = \frac{p}{2}(1-q)\left[\frac{1+q}{1-q} - j\cot \omega T/2\right] \tag{4-3}$$

This is written in a more useful form by applying Eqs (3-6) to p and q and noting that

$$\frac{\omega T}{2} = \frac{\omega}{2\pi/T}\pi = \frac{\omega}{\omega_s}\pi = r\pi \tag{4-4}$$

The final result is then

$$G^*(j\omega) = \frac{\lambda T}{2}[(2\gamma-1) - j\cot r\pi] \tag{4-5}$$

The magnitude and argument of $G^*(j\omega)$ are readily found from Eq (4-5) according to their definitions in Eqs (2-7) and (2-8). The appropriate relations are

$$M = \frac{\lambda T}{2}[(2\gamma-1)^2 + \cot^2 r\pi]^{\frac{1}{2}} \tag{4-6}$$

$$\phi = -\tan^{-1}\left\{\frac{\cot r\pi}{2\gamma-1}\right\} \tag{4-7}$$

These last three relations tell us much about the properties of the ZOH-TI, properties which [6] shows apply generally to all of the tunable integrators developed to date. One general observation is that the phase characteristics of the ZOH-TI are independent of λ, and depend on T only through r. The performance of a numerical integrator depends not on the frequency of the input signal nor the length of the integration step, but upon the relative size of that step compared to that frequency. A second observation is that λ and T will have a scaling effect on the output magnitude. Thus, we have a simple analytic justification for Smith's observation that γ affects the transient response and λ. the steady state solution.[4]

There are two types of symmetry apparent. One is the symmetry of M as a function of γ about some value γ_M; here $\gamma_M = 0.5$. Thus, the Bode magnitude

plot for a rectangular integrator (γ=1) is the same as for an Euler integrator (γ=0). Smith observes that this arises naturally for the ZOH-TI because "the zero-order hold integrand reconstruction process introduces exactly a half period of lag which is compensated by a half period of lead when $\gamma = +\frac{1}{2}$."[15]

The second type of symmetry is the mirror imagery of ϕ as a function of r. There is often, though not always, a γ_ϕ about which equal displacements of γ produce equal displacements of ϕ. For the ZOH-TI the value of γ_ϕ is the same as γ_M (i.e., 0.5) and $\phi(\gamma_\phi) = -90^o$ for all r.

The last general property is the appearance of the term cot $r\pi$ in all of the gain relations. As r becomes small, this term dominates the expression for M and all integrators have approximately the same magnitude characteristics. If we expand the expression for M with $r << 1$, we find that $M \approx \lambda/\omega$ and we match the true integrator gain slope of 20 db/decade. Since the phase of the integrator approaches -90^o for all values of λ when r is small, we have a very cogent demonstration of why small integration steps result in satisfactory performance by nearly any integrator.

The Bode plots of M_{db} and ϕ versus $\log_{10} r$ shown in Figures 16 and 17, respectively, were generated using T = 0.002 sec. Due to information limitations of the sampling theorem, only values of $r \leq \frac{1}{2}$ have been used. The symmetry of M and mirror imagery of ϕ are evident, as is the 20 db/decade slope of M for small r. The slope remains at that value longest for γ = 0.2 and 0.8, these then being the best choices for controlling amplitude distortion. It is interesting to note that at $r = \frac{1}{2}$ (log r = -0.3) the output of the trapazoidal integrator would be zero (since the integrand samples are being taken precisely one-half period apart). We observed in Chapter 3 the difference in output of the second order system when there was a lot of lead compensation (γ=1) and when there was none (γ=0), the latter case being unstable at large T (or r). Both situations produce the same magnitude characteristics, but greatly different phase characteristics, for the open-

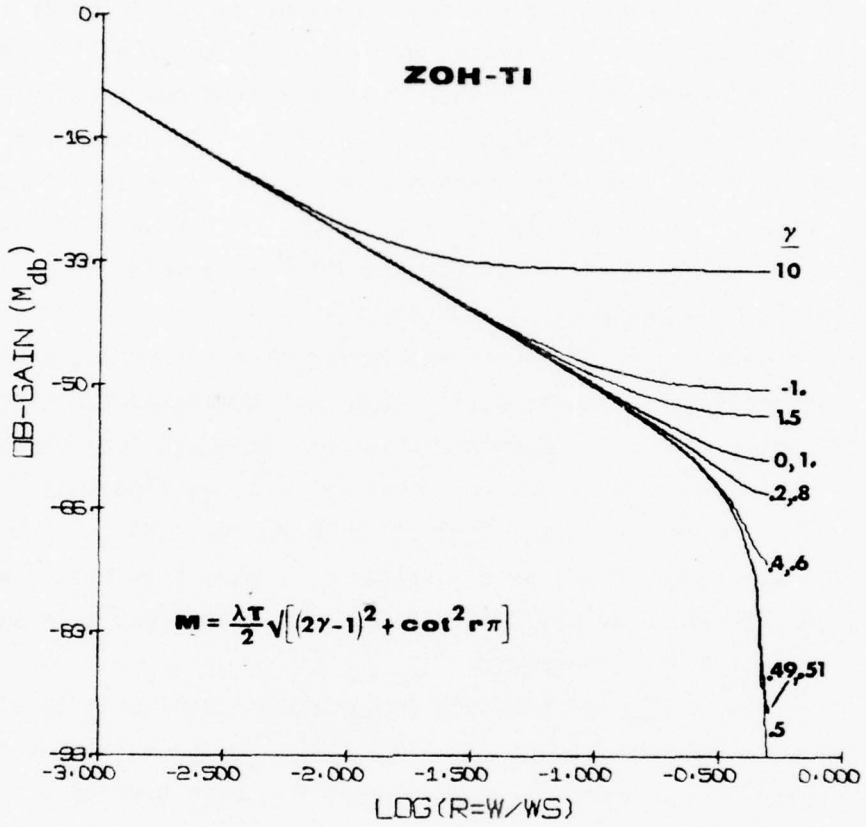


Figure 16. ZOH-TI Bode Magnitude Plot

loop integrator. Since both gain margin and phase margin[†] affect system stability, it is reasonable to say that the phase characteristics are more important to integrator performance. This was empirically observed by Smith[4] and is here analytically explained.

[†]Gain and phase margin are defined by the relations

$$\text{Gain Margin} = -|G(j\omega_\pi)|_{db} \text{ for Arg } \{G(j\omega_\pi)\} = -180^\circ$$

$$\text{Phase Margin} = 180^\circ + \text{Arg } \{G(j\omega_1)\} \text{ for } |G(j\omega_1)| = 1$$

These terms are normally applied to the open-loop transfer function, GH, of a closed-loop control system.

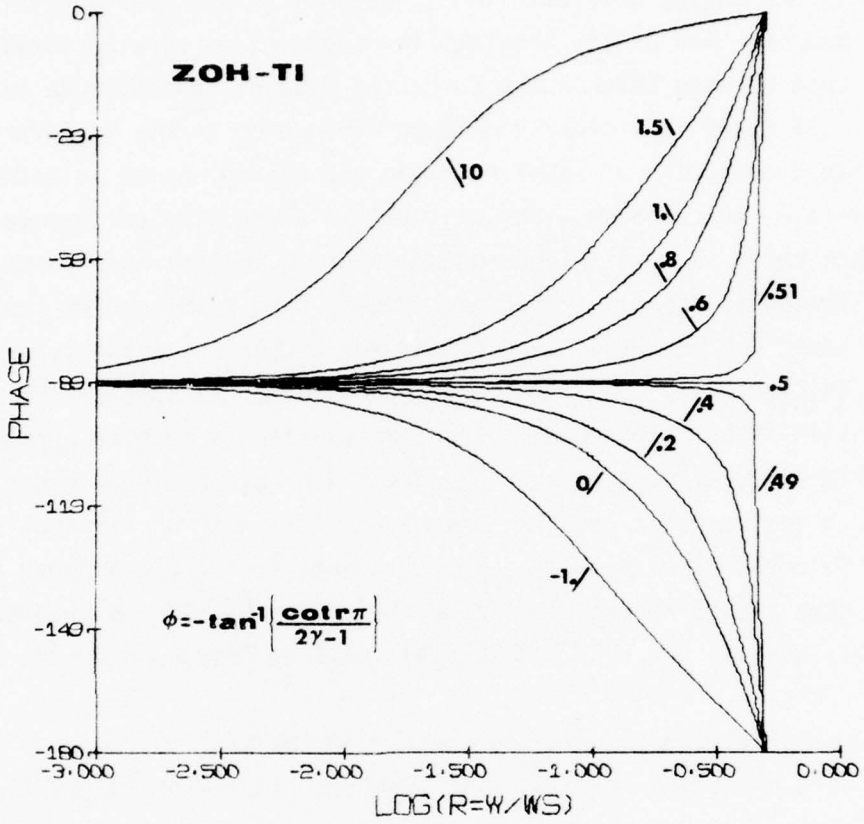


Figure 17. ZOH-TI Bode Phase-Angle Plot

Figure 17 clearly shows the lead-compensation effect of γ for $\gamma > \frac{1}{2}$ and the additional lag (beyond -90^0) when $\gamma < \frac{1}{2}$. Just as was observed in Figure 13, increasing the integration step, and concomitantly the truncation error, does not necessarily induce greater lag and instability, though it does induce error. The best tuning to control phase distortion is $\gamma = \frac{1}{2}$, for then $\phi \equiv -90^0$. Combining ideal phase characteristics and relatively good magnitude characteristics, the trapazoidal integrator is obviously, and well-known to be, a good integrator. It is interesting to note that except for $\gamma = 0.5$, all of the phase curves pass through 0^0 or -180^0 of phase at $r = \frac{1}{2}$.

4.4 Tuning Based Upon the Frequency Response

The frequency response data for the ZOH-TI have been related after-the-fact to the previously discussed results of the second-order system forced at

resonance. One would, however, like to be able to use this analysis in an a priori manner. For simple problems the approach is straightforward. A tradeoff must be made between the competing demands of amplitude and phase fidelity. If there is a clearly dominant frequency in the integrand, then the appropriate combination of sampling ratio and tuning can be selected to meet the criteria of the problem, such as total solution time or computer cycle time. When there are multiple frequencies of equivalent importance, the difficulty of tuning is greatly increased. When a single integrator cannot be tuned to cover the frequency band of interest, then one possible approach would be to spectrally separate the integrand prior to integration with recombination after integration. The complexities of such an approach and the difficulties imposed by nonlinear problems have not been examined by this author. In very complex problems one may be left with no alternative but to tune the integrator empirically. In any event, frequency response data such as shown here should not be used alone, but in combination with other analytical tools such as the root locus. This will be discussed in the next chapter.

Return briefly to our example problem of the previous chapter to see how we might have proceeded in an a priori manner. If the objective is to maximize the step size with phase error the primary criterion, then one would choose γ near 0.5. Making γ a little larger than 0.5 gives a cushion of phase lead and also improves the magnitude characteristics. The values of $r = \omega_d/\omega_s$ for the four time steps to be considered are 0.95×10^{-3}, 0.95×10^{-2}, 0.47×10^{-1}, and 0.95×10^{-1} respectively for T = 0.001, 0.010, 0.050, and 0.100. The logarithm of the largest value is -1.02. Looking at Figures 16 and 17, one can see that a value of γ only a little larger than 0.5 will provide nearly ideal magnitude and phase characteristics. Thus, one should expect that $\gamma = 0.52$ will allow use of T = 0.100. The value of 0.52 was used to generate the step response of Figure 18. Compare this figure with Figure 15, and note the increase in range of allowable step sizes that is afforded by proper tuning. Clearly, frequency response data can aid in the selection of the integration formula to be used in the solution of differential equations.

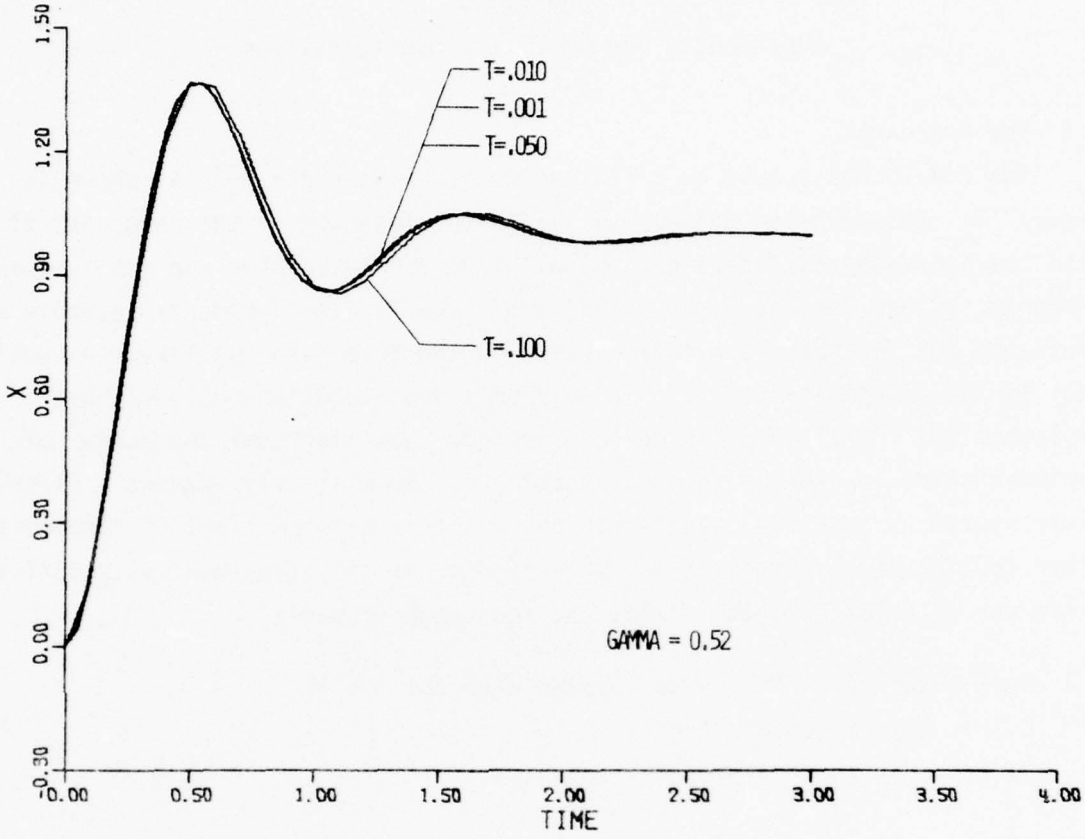


Figure 18. ZOH-TI Step Response with $\gamma = 0.52$

CHAPTER 5
Root-Locus and Stability Considerations

5.1 The Approach

The ZOH-TI has a pole at $z = 1$ and a zero at $z = q = \frac{\gamma-1}{\gamma}$, as shown in Figure 10. Changing the value of γ shifts the location of the zero, and it also has a pronounced impact on the output of the integrator and its frequency response, as was demonstrated in Chapters 3 and 4. The frequency response was evaluated for the open-loop integrator, and the pole-zero map likewise applied only to the integrator itself. To perform a root-locus analysis we must implement the ZOH-TI in a closed-loop system. How important the manner of implementation is will be shown in Chapter 6. We will only examine a first-order system so that the details can be kept to a minimum level of complexity. After looking at this root locus, we will discuss the assessment of stability using the root-locus together with the frequency response.

5.2 Root Locus of a First-Order System with the ZOH-TI

5.2.1 The Implementation

We consider the differential equation

$$\tau\dot{x} + x = f(t) \tag{5-1}$$

where τ is the system time constant. The block diagram of this simple first-order system is given in Figure 19.

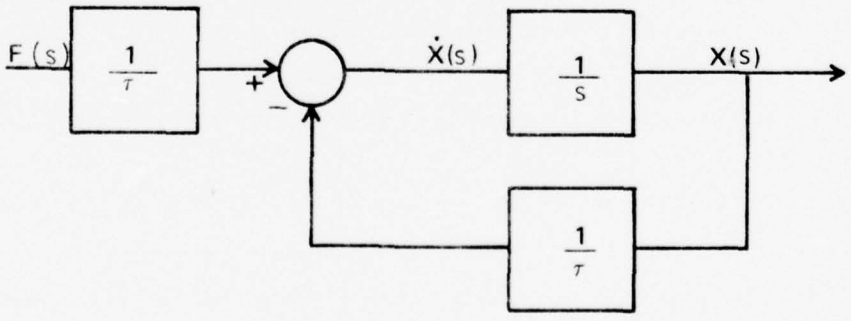


Figure 19. First-Order System Block Diagram†

†In the diagram we could remove the $1/\tau$ blocks from the input and feedback paths by replacing $1/s$ with $1/\tau s$. This form of simplification will be done in subsequent block diagrams.

We have seen that the ZOH-TI is an implicit algorithm, requiring some type of predictor for x_{n+1} or extrapolator for $\dot{x}_{n+1}$; we shall use an Euler predictor in combination with the ZOH-TI corrector. The section of computer code dealing with the actual integration of $\dot{x}$ could be programmed in FORTRAN as the following sequence of steps:

$$\left.\begin{aligned} Y &= F/TAU \\ DX1 &= Y-X/TAU \\ X1 &= X+T*DX1 \end{aligned}\right\} \qquad (5\text{-}2)^{\dagger}$$

$$\left.\begin{aligned} DX &= Y-X1/TAU \\ X &= X+LAMDA*T*\left(GAMMA*DX+(1-GAMMA)*DX1\right) \end{aligned}\right\} \qquad (5\text{-}3)$$

Equations (5-2) represent the Euler predictor and Eqs (5-3), the ZOH-TI predictor. We assume that F is evaluated prior to Eqs (5-2) and that the appropriate iterative loops are defined.

Let us look at these equations as recursion relations and attempt to follow the flow of information through the software.[††] Neglect for the moment the association of any time base with the variables, since we separately track the independent variable on the digital computer by incrementing it with the integration step. Suppose that we are in the 100th pass through Eqs (5-2) and (5-3), having completed 99 passes. Subscripting the variables to indicate the number of the respective evaluation, we can now write

$$\left.\begin{aligned} Y_{100} &= F_{100}/TAU \\ DX1_{100} &= Y_{100}-X_{99}/TAU \\ X1_{100} &= X_{99}+T*DX1_{100} \\ DX_{100} &= Y_{100}-X1_{100}/TAU \\ X_{100} &= X_{99}+LAMDA*T*\left(GAMMA*DX_{100}+(1-GAMMA)*DX1_{100}\right) \end{aligned}\right\} \qquad (5\text{-}4)$$

[†] We could more simply write DX1 = (F-X)/TAU, but have chosen this form to relate to the subsequent discussion on implementation.

[††] The idea for this approach was provided by J.M. Smith in private conversation.

Note that X_{99} is the stored value of X until the last computation.

At this time, associate a time base with X and, recalling that $z^{-1}X_{n+1} = X_n$, identify the increment T in time with a unit increment in the subscript. Thus, we can transform Eqs (5-4) into algebraic form:

$$\left.\begin{aligned} Y(z) &= F(z)/\tau \\ DX1(z) &= Y(z)-z^{-1}X(z)/\tau \\ X1(z) &= z^{-1}X(z)+T\cdot DX1(z) \\ DX(z) &= Y(z)-X1(z)/\tau \\ X(z) &= z^{-1}X(z)+\lambda T[\gamma DX(z)+(1-\gamma)DX1(z)] \end{aligned}\right\} \qquad (5\text{-}5)$$

5.2.2 The Block Diagram

It is from Eqs (5-5) that we can derive the transfer function of the software algorithm and then perform the root-locus analysis.

The question at this juncture is what is the appropriate transfer function for the integrator? Is it X/DX or is it X/DX1? Since X1 and DX are internal to the algorithm, we should seek X/DX1. From the last of Eqs (5-5), we have

$$\frac{X(z)}{DX1(z)} = \frac{\lambda T(1-\gamma)z}{z-1} + \frac{\lambda T\gamma z}{z-1}\,\frac{DX(z)}{DX1(z)} \qquad (5\text{-}6)$$

We can obtain DX/DX1 by subtracting the second and fourth of Eqs (5-5) and then substituting for X1 from the third. Proceeding, we have

$$DX1(z)-DX(z) = \frac{1}{\tau}\,[z^{-1}X(z)+T\cdot DX1(z)-z^{-1}X(z)] = \frac{T}{\tau}\,DX1(z)$$

or

$$\frac{DX(z)}{DX1(z)} = 1-T/\tau \qquad (5\text{-}7)$$

Combining Eqs (5-6) and (5-7) we get the desired relation

$$\frac{X(z)}{DX1(z)} = \lambda T(1-\gamma T/\tau)\,\frac{z}{z-1} \qquad (5\text{-}8)$$

This same result could be obtained from the methods of block-diagram manipula-

tion, or signal-flow graph analysis, by drawing the diagram corresponding to Eq (5-5) and then reducing it to simple form.

Equation (5-8) is the "integrator" transfer function. The integrator we have, however, is a result of the manner in which we implemented (or programmed, if you will) the algorithm. We did not simply substitute Eq (3-5) for the ZOH-TI and T/(z-1) for the Euler integrator into the 1/s block of Figure 19. In fact, we did not even program a true Euler integrator as can be seen from Eq (5-5) and will be made very clear in the next chapter.

In this formulation we note that the position of the open-loop zero is no longer a function of the value of γ. The tuning parameters, as well as the ratio T/τ (which is analogous to our sampling ratio $r = \omega/\omega_s$ in the frequency response), now affect the forward loop gain. As this gain varies, the pole location will change, thereby tracing out the root locus.

The integrator of Eq (5-8) can now be incorporated in the overall software system by noting the inputs to DX1 and the feedback of X given by Eq (5-5). We will see from the result that this will not be equivalent to simply substituting Eq (5-8) for 1/s in Figure 19. The inputs to DX1(z) are $\frac{1}{\tau} F(z)$ and X(z) fed back through z^{-1}/τ. As noted for Figure 19, the $1/\tau$ factor can be moved to the forward-path side of the summing junction. With this simplification we have the block diagram of Figure 20.

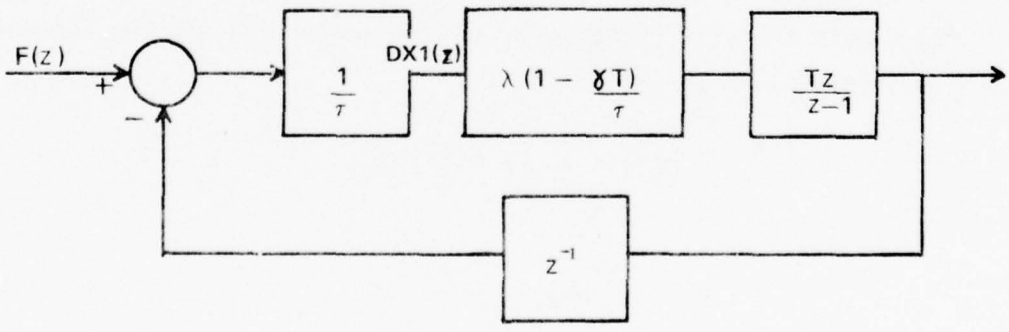


Figure 20. ZOH-TI/Euler Implementation of $\tau\dot{x} + x = f$

Comparison of Figures 19 and 20 shows that the key difference from direct substitution of Eq (5-8) for 1/s is the delay in the feedback loop that is due to the Euler predictor. This delay has a significant impact on stability.

Another interesting comparison shows the advantage of using the ZOH-TI for the corrector in this analysis. If we retrace our steps using a rectangular integrator instead of the ZOH-TI, then the block diagram of Figure 20 takes the form shown in Figure 21.

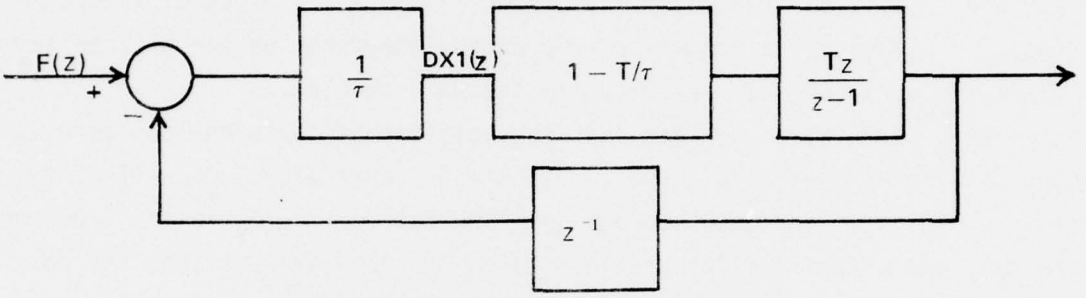


Figure 21. Rectangular/Euler Implementation of $\tau\dot{x} + x = f$

What the ZOH-TI provides is an ability, other than by means of step size, to control the forward loop gain and to thereby control overall system stability. Note that in both diagrams the block Tz/(z-1) appears: this is the open-loop transfer function of the rectangular integrator.

5.2.3 The Root Locus

We initially need to compute the open- and closed-loop transfer functions. If the forward path transfer function is designated as G(z) and the feedback path as H(z), then Figure 20 tells us that

$$G(z) = \frac{\lambda T}{\tau}\left(1 - \frac{\gamma T}{\tau}\right)\frac{z}{z-1}$$

and

$$H(z) = z^{-1}$$

The open-loop transfer function is the product of these:

$$GH \triangleq G(z)H(z) = \lambda\left(1 - \frac{\gamma T}{\tau}\right)\frac{T/\tau}{z-1} \qquad (5\text{-}9)$$

The complete, closed-loop transfer function $X(z)/F(z)$ is given by the relation[†]

$$\frac{X(z)}{F(z)} \triangleq T_{CL}(z) = \frac{G}{1+GH} = \frac{Kz}{z-(1-K)} \tag{5-10}$$

where

$$K = \frac{\lambda T}{\tau}\left(1 - \frac{\gamma T}{\tau}\right) \tag{5-11}$$

Equation (5-10) is the transfer function of the computer software that solves the differential equation $\tau\dot{x} + x = f$ with an Euler predictor and a ZOH-TI corrector. One could in fact very simply replace the entire algorithm of Eq (5-2) with the single difference equation corresponding to Eq (5-10); i.e.,

$$x_{n+1} = (1-K)x_n + Kf_{n+1} \tag{5-12}$$

Taking $x(t=0) = 0$, $f(t) = u(t)$, $\tau = 1$ sec, $T = 0.1$ sec, $\lambda = 1$, and $\gamma = 0.5$, we find the values of x at $t = 0.1$ sec and $t = 0.2$ sec to be respectively 0.095 and 0.180975 by both Eqs (5-2) and (5-12). The difficulty (or perhaps advantage, depending on one's perspective) with Eq (5-12) is the need to predict the forcing function $f(t)$ rather than the state x at t_{n+1}.

The root locus maps the change in pole location as the forward-path gain increases. Beginning at the open-loop position the poles move toward the open-loop zeros. Equations (5-9) and (5-10) show a single open loop pole at $z = 1$[††] and a closed-loop zero at $z = 0$. Thus, the single pole takes on the value $1-K$, and as K increases it moves from the point $(1,0)$ to $-\infty$ along the $\mathrm{Re}\{z\}$ axis. When $K = 2$, the pole is at $(-1,0)$. For $\gamma T/\tau$ sufficiently large we could make $K < 0$, but that would immediately make the system unstable. Hence, we plot the root locus as shown in Figure 22.

[†]See [16], or any other similar text, for a discussion of sampled-data control system analysis.

[††]Note that we are no longer designing a simple integrator, so that this pole at $z = 1$ is not necessarily ideal.

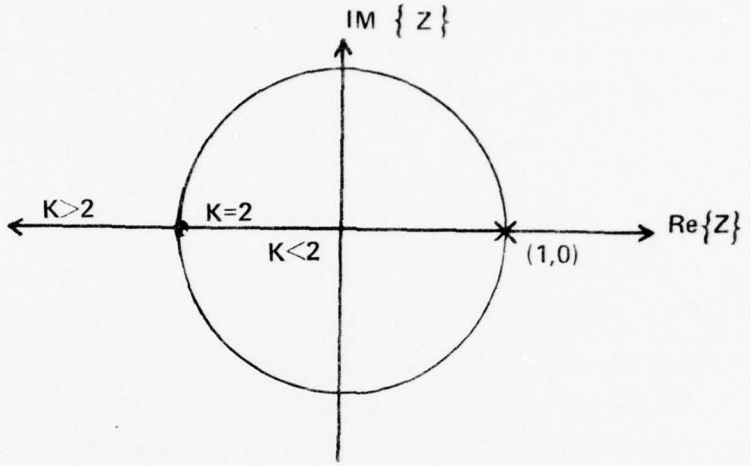


Figure 22. Root Locus for ZOH-TI Euler Implementation

The root locus can now be used for two important analyses. The first is the matching of the pole of this discrete model to that of the continuous system. This type of analysis is performed in [14]. Our attention in this report will be restricted to the second use of the root locus: stability analysis.

5.3 Stability Considerations

5.3.1 General Comments

Recall from the frequency-response analysis that stability does not guarantee accuracy, and thus we see how to use the frequency-response analysis along with root-locus analysis. Fidelity of the integration can be specified by examination of the former and stability of the result can be ascertained from the latter. One can also combine these two analyses with root-matching techniques, and then trade off the competing requirements.

Before proceeding, note that stability information is also available from the Bode plots of Chapter 4. The gain and phase margins both being greater than zero is the indicator of stability, where

$$\left.\begin{aligned} \text{Gain Margin} &= -|G(j\omega_\pi)|_{db} \text{ for Arg } \{G(j\omega_\pi)\} = -180^o \\ \text{Phase Margin} &= 180^o + \text{Arg } \{G(j\omega_1)\} \text{ for } |G(j\omega_1)| = 1 \end{aligned}\right\} \quad (5\text{-}13)$$

From Figures 16 and 17, it is apparent that these quantities are not necessarily readily found for the simple integrator. In these figures, ω_1 corresponds to log r = -3.5, for which the phase margin is $\approx +90^0$ regardless of γ. The gain margin, however, does not exist for $\gamma \geq 0.5$. For $\gamma < 0.5$ it is the negative value of the gain, in db, at r = 0.5. These data alone would tend to indicate that no stability problems should occur for any value of γ shown. The integrator, however, is part of an overall system and its lag, contributing to the system's lag, can produce instability. Note also that T just scales the gain in Eq (4-6), thereby only serving to shift the magnitude plots along the ordinate. As T increases, the curves shift upward and cause both gain and phase margins to decrease when $\gamma < 0.5$, thereby decreasing stability. When $\gamma > 0.5$, the gain margin decreases but the phase margin increases. Thus, we have yet another bit of insight as to what the impact of changing step-size is, insight with a different perspective than that of truncation error. We also see why the frequency-response data should be used in conjunction with other analyses such as the root-locus, since the former is not conclusive.

5.3.2 Analysis of the First-Order System

The region of stability in the z-plane lies within the unit circle, and therefore the sampled-data system represented by Eq (5-10) (which is our software transfer function) is stable for values of K in the interval [0,2]. This interval presents us with a range of possible values for the tuning parameters λ and γ and the sampling ratio T/τ, all of which will yield system stability.

To see what choices of γ might be available to us, let $\lambda = 1$; then the limits on K give

$$0 \leq K \leq 2$$

$$0 \leq \frac{T}{\tau}\left(1 - \frac{\gamma T}{\tau}\right) \leq 2$$

$$\frac{T}{\tau} \geq \gamma \geq \frac{\tau}{T}\left(1 - \frac{2\tau}{T}\right) \qquad (5\text{-}14)$$

A plot of acceptable values of γ versus T/τ then takes the form of Figure 23.

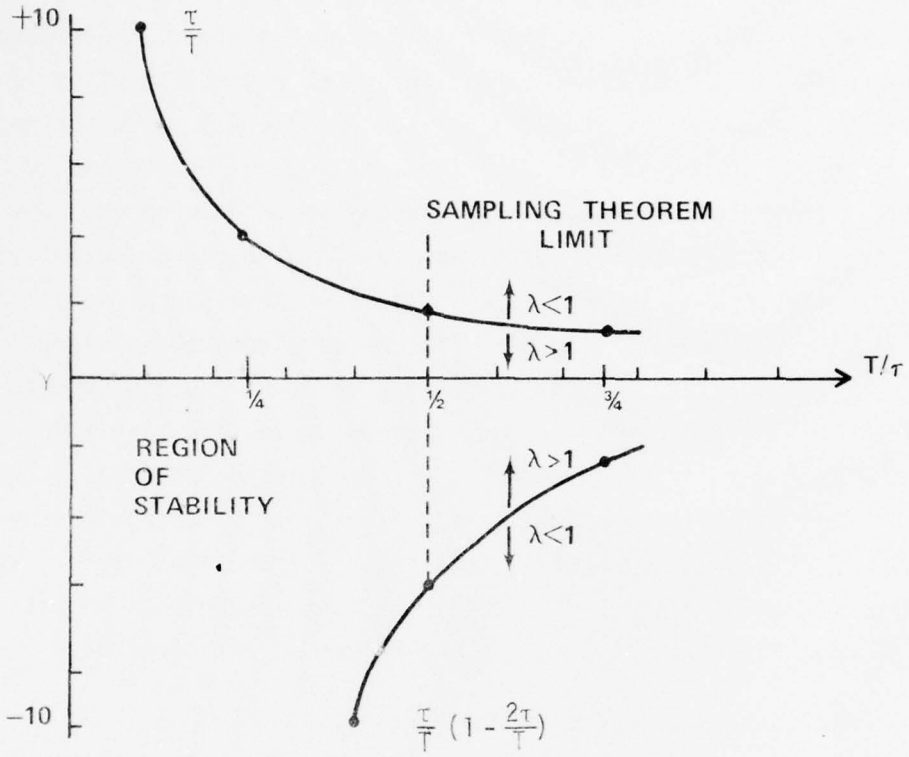


Figure 23. Stability Region for Tuning

It is interesting to note that the lower bound on K generates the upper bound on γ and vice versa. This is quite logical since larger values of K move the system towards instability and lower values of γ move the integrator towards instability. This figure also parallels the results of Chapter 3 in that it points out that for small values of T/τ, the choice of γ has much latitude. (See, for example, Figures 12-14.)

5.3.3 A Trial Solution

We have looked at the frequency response of the ZOH-TI, and we have looked at the root locus of the ZOH-TI/Euler pair applied to a first order system and the stability implications of that application. A brief example demonstrates one possible approach to tuning based on those analyses. Let τ = 1 sec, f = cos t (ω = 1 rad/sec) and choose an integrator with as large a time step as possible to give reasonable fidelity. Figures 16 and 17 show that for γ in

the range .49 to .51 we obtain a fairly good magnitude characteristics up to about log $r = -1$ or $r = 0.1$. Figure 23 shows this should be well within the region of stability of the system, and that we should be able to approach the information limit of $T/\tau = \frac{1}{2}$. The value of r would imply $T = \frac{2r\pi}{\omega} = 0.63$, and $T/\tau = \frac{1}{2}$ would imply $T = 0.5$.[†] Let's be conservative and choose $T = 0.2$, which gives $r = 0.032$ and $T/\tau = 0.2$. From Figure 16, we see that M_{db} for $r = 0.032$ (log $r \approx -1.5$) is about -40 db. Since that figure was produced for $T = .002$ rather than $T = 0.2$, we raise the curve 40 db[††] and see that λ should be about unity (since $1/\omega = 1$ gives the true integrator gain as zero db and this is what we have).

The true solution to Eq (5-1) when $f(t) = \cos \omega t$ is

$$x_t(t) = \frac{-\tau}{1+(\omega\tau)^2} e^{-t/\tau} + \frac{1}{1+(\omega\tau)^2} (\cos \omega t + \omega\tau \sin \omega t) \qquad (5\text{-}15)$$

Using λ slightly larger than unity, $\lambda = 1.01$, and $\gamma = 0.51$, the computed solution $x_c(nT)$ compares to the true solution as shown in Table 5. The value of the gain K is 0.18 from Eq (5-11). Though not providing excellent accuracy, this simple combination of the Euler and ZOH-TI has given a reasonable reproduction of the true solution. Better tuning and a slightly smaller time step would improve the results. Accuracy was not the point of this demonstration, however. The objective was to show how one could initiate the tuning process by using the frequency-response and root-locus data.

[†]Note that one criteria reflects sampling relative to the input f(t) and the other relative to the system time constant τ. One should not expect two such similar values of T to arise in general.

[††]Since T scales the magnitude, as shown by Eq (4-6), the curve for $T = 0.2$ is simply 40 db above that for $T = 0.002$.

Table 5: True Solution vs ZOH-TI/Euler

t	True $x_t(t=nT)$	ZOH-TI/Euler $x_c(nT)$
0.8	.482	.458
1.0	.507	.473
1.2	.497	.453
3.8	-.713	-.705
4.0	-.714	-.720
4.2	-.688	-.708
6.8	.681	.698
7.0	.705	.708
7.2	.701	.690
10.0	-.692	-.704
10.2	-.707	-.706
10.4	-.694	-.680

I would make one more point before proceeding. If the z-transform of cos ωt is multiplied by $T_{CL}(z)$ from Eq (5-10) and the result is expanded by partial fractions, one of the terms will have the factor $\frac{z}{z-(1-K)}$. The solution component corresponding to this term has the form $C[1-K]^{n+1}$, where C is an explicit function of K, T and ω. This is the term corresponding to the exponential in Eq (5-15), and when $K > 2$ it will oscillate with increasing size, corresponding to the previously identified instability for values of K greater than two.

5.3.4 Stability from Two Perspectives

When most numerical analysts talk about the stability of a numerical procedure, they refer to the propagation of errors through the algorithm and question whether or not an error introduced at one point will die out or grow without bound. Rosko presents a clear presentation of this type of analysis in [12].

Following Rosko's presentation, take the multistep method

$$x_{n+1} = \sum_{i=0}^{m} a_i x_{n-i} + T \sum_{i=-1}^{m} b_i \dot{x}_{n-i} + e_{R_n} \qquad (5\text{-}16)$$

where e_{R_n} is the roundoff error at the n^{th} step. A similar relation holds for the true value of x at t_{n+1}, when e_{R_n} is replaced by e_{T_n}, the local truncation error. The difference between the two values of x is the error, and it propagates according to the relation

$$e_{n+1} = \sum_{i=0}^{m} a_i e_{n-i} + T \sum_{i=1}^{m} b_i \dot{e}_{n-i} + E_n$$

where $E_n = e_{T_n} - e_{R_n}$. Given that $\dot{x} = f(x,t)$, the mean-value theorem allow us to write this relation as

$$\left(1 - Tb_{-1} f_{x_{n+1}}\right) e_{n+1} = \sum_{i=0}^{m} \left(a_i + Tb_i f_{x_{n-i}}\right) e_{n-i} + E_n \qquad (5\text{-}17)$$

where

$$f_{x_n} = \frac{\partial f}{\partial x}(\xi_n, t_n) \quad \text{and} \quad \xi_n \in [x_n, x_t(t_n)]$$

Typically, the assumption is made that f_x is at worst slowly varying, so it is taken as a constant, f_x. The characteristic equation corresponding to Eq (5-17) is then, in terms of z,

$$0 = \psi(z) = \left(1 - Tb_{-1} f_x\right) z^{m+1} - \sum_{i=0}^{m} \left(a_i + Tb_i f_x\right) z^{m-i} \qquad (5\text{-}18)$$

To be stable we want all the roots of Eq (5-18) to lie within the unit circle. Consider, for example, an Euler integrator. The relation $x_{n+1} = x_n + T\dot{x}_n$ gives us from Eq (5-16)

$$m = 0 \text{ (1-step integrator)}$$

$$a_0 = 1$$

$$b_{-1} = 0$$

$$b_0 = 1$$

Equation (5-18) becomes

$$0 = \psi_1(z) = z - (1 + Tf_x) \tag{5-19}$$

and we have the criterion for stability as

$$|1 + Tf_x| < 1 \tag{5-20}$$

For f_x real, we have the region of stability given by the relation

$$Tf_x \in [-2,0] \tag{5-21}$$

Compare this result with one obtained via the type of analysis presented earlier in this chapter. For the first-order system examined, as we will see in Chapter 6, a true Euler integrator can be implemented to provide the block diagram of Figure 24.

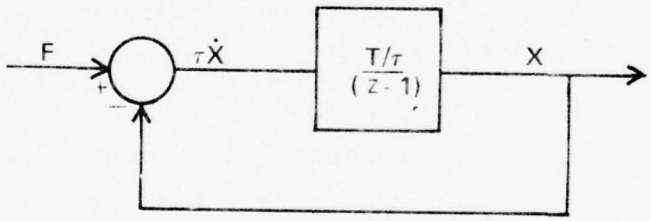


Figure 24. True Euler Implementation of $\tau\dot{x} + x = f$

The closed-loop transfer function is

$$T_{CL}(z) \approx \frac{T/\tau}{z-(1-T/\tau)} \tag{5-22}$$

This is stable for

$$\left.\begin{aligned} |(1-T/\tau)| &< 1 \\ 0 < T/\tau &< 2 \end{aligned}\right\} \tag{5-23}$$

These relations are identical to Eqs (5-20) and (5-21), for $f_x = -1/\tau$.

Now consider Simpson's integrator (see Eq (2-4)), in the form

$$x_{n+1} = x_{n-1} + \frac{T}{3}\left(\dot{x}_{n+1} + 4\dot{x}_n + \dot{x}_{n-1}\right) \tag{5-24}$$

The parameters of Eq (5-16) are

$$m = 1 \text{ (two-step integrator)}$$

$$a_0 = 0,\ a_1 = 1$$

$$b_{-1} = 1/3,\ b_0 = 4/3,\ b_1 = 1/3$$

and the characteristic equation is

$$0 = \psi_2(z) = (1-T/3\ f_x)\ z^2 - 4T/3\ f_x z - (1+T/3\ f_x) \tag{5-25}$$

The roots of this quadratic are

$$z = \frac{2Tf_x \pm \sqrt{9+3(Tf_x)^2}}{3-Tf_x}$$

Rosko notes that for all real values of Tf_x one or the other of these two roots will be outside the unit circle, except for $Tf_x = 0$ which puts the two roots on the unit circle. Thus, as we pointed out in Chapter 2, this integrator is not suitable for solving differential equations.

If we look at the basic integrator transfer function, as was done in Chapter 4, then for Simpson's integrator we have

$$G(z) = \frac{T/3\ (z^2 + 4z + 1)}{z^2 - 1}$$

The poles of the integrator itself, without regard to the system it is employed in, are at $z = \pm 1$, and the zero outside the unit circle at $-2-\sqrt{3}$ will attract a pole whenever the integrator is implemented in the solution of a differential equation, clearly an unstable situation.

Thus, in these two cases at least, looking at either the error equation or looking at the basic integrator transfer function or the overall software transfer function has led us to the same conclusion.

Observe that for both the Euler and the Simpson integrators, the characteristic polynomials of the basic integrator are the same as for the respective error equation, Eqs (5-19) and (5-25), when the dynamics are neglected; i.e.,

$$\text{Euler} \qquad \psi_1(z)\big|_{Tf_x=0} = z-1$$

$$\text{Simpson} \qquad \psi_1(z)\big|_{Tf_x=0} = z^2-1$$

The relation between what has been done in this report and some of the more familiar approaches to stability is now a little clearer. Looking at the basic integrator pole locations gives a quick look at stability independent of system dynamics. A feel for what will happen at relatively small time steps is gained, for then $Tf_x << 1$.

The frequency response data are based upon sampling ratios relative to the input signal f. The root locus analysis, and the error propagation analysis, are dependent upon sampling relative to the characteristic system frequencies (recall the importance of T/τ in our analysis of this chapter). Therefore, we have formulated in these last three chapters an initial capability to design, using control-theory methods, an integrator to meet criteria for sampling relative to dominant frequencies in both the driving function and the system itself. The approach is not yet fully or rigorously defined, but a beginning has been made.

CHAPTER 6
The Effect of Implementation

6.1 Why Worry?

One might easily be tempted to take a given numerical algorithm and program it for the computer in any convenient manner without regard to the manner of programming. That programming, however, has a significant impact on the performance of the software and can make the difference between a successful solution and an unsuccessful one. Shampine, et. al.[2] noted this phenomena with respect to the GEAR and DIFSUB routines which use essentially the same numerical methods.

To fully analyze a package as large as GEAR or DIFSUB in the manner of this chapter would be a tremendous undertaking, and based upon the success of those routines would not be justified now. Future developments, however, should pay attention, initially, to the manner of implementation. The methods of control theory provide a convenient means for the analysis, for separable parts of the software can be designed and their subsequent interactions as coupled systems can be systematically analyzed. The emphasis of this chapter is on an approach to analyzing numerical methods rather than a specific method, so I will only present an elementary example of the considerations involved.

6.2 Implementation of $\tau\dot{x} + x = f$

Let us take the same example problem as was used in Chapter 5, but deal only with the predictor part of the method employed therein, the Euler integrator. The way this integrator was implemented in Chapter 5 did not produce a true Euler integrator. The difference between that implementation and true Euler implementation is cogently demonstrated by a frequency response analysis such as the one in Chapter 3.

6.2.1 A "Mixed-Mode" Euler Integrator

For convenience, we repeat Eq (5-2) at this point, but change DX1 and X1 to DX and X, respectively.

$$\left.\begin{aligned} Y &= F/TAU \\ DX &= Y-X/TAU \\ X &= X+T*DX \end{aligned}\right\} \qquad (6\text{-}1)$$

As noted in the previous chapter, we assume that F is evaluated prior to the computation of Y. It is also assumed that the independent variable is updated prior to computing F. Though this looks like an Euler integrator, it is not a true Euler integrator because of the way in which DX is computed.

Based upon the first part of Eq (5-4) we have the pass-count evaluation (where TI is the independent variable time) as follows:

$$\left.\begin{aligned} TI_{100} &= TI_{99} + T \\ F_{100} &= F(TI_{100}) \\ Y_{100} &= F_{100}/TAU \\ DX_{100} &= Y_{100} - X_{99}/TAU \\ X_{100} &= X_{99} + T*DX_{100} \end{aligned}\right\} \qquad (6\text{-}2)^{\dagger}$$

To compute DX_{100} we have used Y from pass 100 and X from pass 99. Hence, DX_{100} is an approximation to $\dot{x}$ at some time which is correctly associated with neither TI_{99} nor TI_{100}.

In terms of z, we convert Eqs (6-2) to the transformed relations

$$\left.\begin{aligned} Y(z) &= F(z)/\tau \\ DX(z) &= Y(z) - z^{-1}X(z)/\tau \\ X(z) &= z^{-1}X(z) + T\cdot DX(z) \end{aligned}\right\} \qquad (6\text{-}3)$$

where we neglect the first two relations for TI and F, their not being a factor in the result. From these relations, we can write the integrator transfer function

---

†This is the actual implementation used for the predictor of Chapter 5.

$$G(z) = \frac{Tz}{z-1} \tag{6-4}$$

This, however, looks just like a rectangular integrator, having the corresponding difference equation

$$X_{n+1} = X_n + TDX_{n+1} \tag{6-5}$$

We programmed what we thought was Euler, got what looks like rectangular, and actually have neither since DX_{n+1} is not really $\dot{x}_{n+1}$, combining, as it does, Y_{100} and X_{99}.

The block diagram for the overall system is similar to Figure 20. The inputs to G(z) and the feedback of X(z) are evident in Eq (6-3). Thus, we have Figure 25.

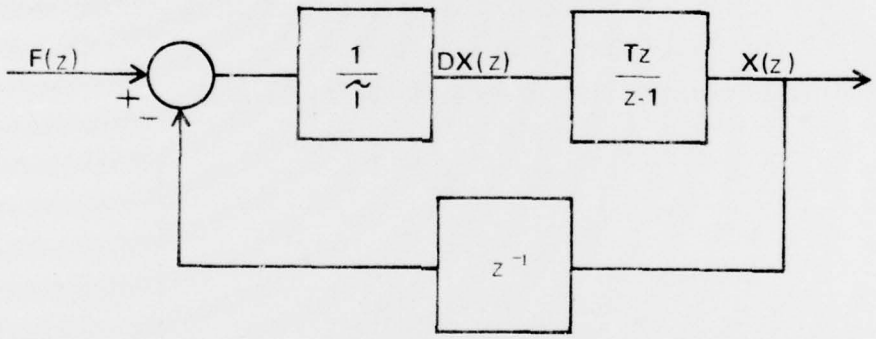


Figure 25. Mixed-Mode Euler Implementation of $\tau\dot{x} + x = f$

The open and closed-loop transfer functions are

$$\left.\begin{aligned} GH &= \frac{T/\tau}{z-1} \\ T_{CL}(z) &= \frac{T/\tau z}{z-(1-T/\tau)} \end{aligned}\right\} \tag{6-6}$$

It is interesting to compare Figure 25 with Figures 20 and 21, and to compare Eqs (6-6) with Eqs (5-9) and (5-10). The forward-path gain is now simply $K = T/\tau$, and the root locus is identical to that shown in Figure 22, with

the different definition of K. Hence, if $K = T/\tau > 2$ the integrator will be unstable. This is precisely the result of Eq (5-21) as obtained from the error propagation equation. It is also the limit for undistorted reconstruction of a signal imposed by the sampling theorem.

6.2.2 A "True" Euler Implementation of $\tau\dot{x} + x = f$

Revise the coding of Eq (6-1) as follows:

$$\left.\begin{aligned} W &= Y \\ Y &= F/TAU \\ DX &= W-X/TAU \\ X &= X+T*DX \end{aligned}\right\} \qquad (6\text{-}7)$$

What we have done is introduce further delay into the system by creation of the dummy variable W. The effect of the additional delay can be seen in the development of the closed-loop transfer function.

Proceeding as before, the pass-count relations are

$$\left.\begin{aligned} TI_{100} &= TI_{99} + T \\ W_{100} &= Y_{99} \\ F_{100} &= F(TI_{100}) \\ Y_{100} &= F_{100}/TAU \\ DX_{100} &= W_{100} - X_{99}/TAU \\ X_{100} &= X_{99} + T*DX_{100} \end{aligned}\right\} \qquad (6\text{-}8)$$

and the important relations in terms of z are

$$\left.\begin{aligned} W(z) &= z^{-1}Y(z) \\ Y(z) &= F(z)/\tau \\ DX(z) &= W(z) - z^{-1}X(z)/\tau \\ X(z) &= z^{-1}X(z) + T*DX(z) \end{aligned}\right\} \qquad (6\text{-}9)$$

The integrator transfer function takes the same form as before:

$$G(z) = \frac{Tz}{z-1} \tag{6-4}$$

We do not have the same result, however, because of what DX_{n+1} represents. It can now be clearly associated with $\dot{x}_n$ because of the delay associated with W.

The inputs to DX and the feedback of X provide the block diagram of Figure 26a, which can be alternatively put in the form of Figure 26b.

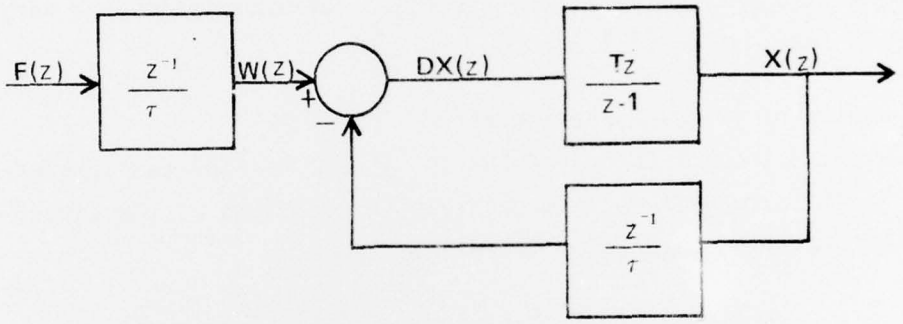


Figure 26a. True-Euler Implementation of $\tau\dot{x} + x = f$

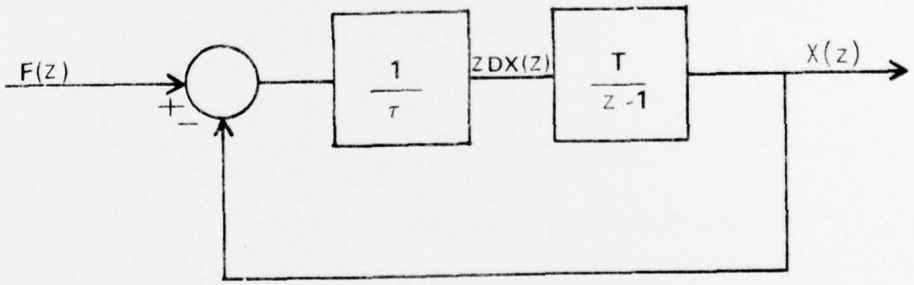


Figure 26b. Reduced Block Diagram

The latter figure is similar to Figure 19 and T/(z-1) is the transfer function for the Euler integrator. This is a rare instance where the desired integrator does simply replace 1/s in the continuous-system block diagram.

Open- and closed-loop transfer functions for the true Euler implementation are

$$\left.\begin{aligned} GH &= \frac{T/\tau}{z-1} \\ T_{CL}(z) &= \frac{T/\tau}{z-(1-T/\tau)} \end{aligned}\right\} \qquad (6\text{-}10)$$

Comparison with Eqs (6-6) shows that the effect of W has been to delete the closed-loop zero (which had been a source of phase lead). The open- and closed-loop pole locations are identical in the two forms of integrator: $z = (1-T/\tau)$. Figure 22 represents the root locus with $K = T/\tau$. Both integrators go unstable at $T/\tau = 2$, but the phase characteristics are very different prior to that, as we will now see.

6.3 Comparison of Frequency Responses

Using the method of Chapter 4 and Eqs (6-6), we find the sampled-data transfer function, magnitude, and phase-angle relations of the closed-loop system with the mixed-mode Euler integrator to be

$$T^*_{CL}(j\omega) = \frac{T/\tau[1-(1-T/\tau)\cos\omega T - j(1-T/\tau)\sin\omega T]}{(T/\tau)^2 + 4(1-T/\tau)\sin^2 T/2}$$

$$\left.\begin{aligned} M_{mm}(r) &= \frac{\frac{1}{2}\,T/\tau}{[(\frac{1}{2}\,T/\tau)^2 + (1-T/\tau)\sin^2 r\pi]^{\frac{1}{2}}} \\ \phi_{mm}(r) &= -\tan^{-1}\left\{\frac{(1-T/\tau)\sin^2 r\pi}{T/\gamma\cos 2r\pi + 2\sin^2 r\pi}\right\} \end{aligned}\right\} \qquad (6\text{-}11)$$

Similarly, for the true-Euler implementation we find

$$T^*_{CL}(j\omega) = \frac{T/\tau[T/\tau - 2\sin^2\omega T/2 - j\sin\omega T]}{(T/\tau)^2 + 4(1-T/\tau)\sin^2\omega T/2}$$

$$\left.\begin{aligned} M_{te}(r) &= \frac{\frac{1}{2}\,T/\tau}{[(\frac{1}{2}\,T/\tau)^2 + (1-T/\tau)\sin^2 r\pi]^{\frac{1}{2}}} \\ \phi_{te}(r) &= -\tan^{-1}\left\{\frac{\sin 2r\pi}{T/\tau - 2\sin^2 r\pi}\right\} \end{aligned}\right\} \qquad (6\text{-}12)$$

Note that Eqs (6-11) apply equally to the ZOH-TI/Euler combination by replacing T/τ with K from Eq (5-11). Also note that $M_{mm} = M_{te}$, which is not surprising, since both implementations have the same open-loop transfer function.

Since we are trying to solve the continuous-system differential equation, the frequency response of the continuous system provides a valid criterion against which to measure the two forms of Euler implementation. The transfer function of interest is

$$T_{CL}(s) = \frac{1}{1 + \tau s}$$

The gain and phase relations for $T_{CL}(j\omega)$ are readily found to be

$$\left.\begin{aligned} M(\omega) &= \frac{1}{(1+\tau^2\omega^2)^{\frac{1}{2}}} = \left[1 + \left(\frac{2\pi r}{T/\tau}\right)^2\right]^{-\frac{1}{2}} \\ \phi(\omega) &= -\tan^{-1}\{\tau\omega\} = -\tan^{-1}\left\{\frac{2\pi r}{T/\tau}\right\} \end{aligned}\right\} \qquad (6\text{-}13)$$

For small values of T/τ and r we should expect to have a close approximation to the continuous system. This is indeed the case, for by using small-angle approximations and neglecting T/τ compared to unity we have from Eqs (6-11) and (6-12)

$$M_{mm} = M_{te} \approx \frac{\frac{1}{2}\, T/\tau}{[(r\pi)^2 + (T/\tau)^2\ \frac{1}{4}]^{\frac{1}{2}}} = \frac{1}{[1 + \tau^2\omega^2]^{\frac{1}{2}}}$$

and

$$\phi_{mm} \approx -\tan^{-1}\left\{\frac{2r\pi}{T/\tau}\right\} = -\tan^{-1}\{\tau\omega\}$$

$$\phi_{te} \approx -\tan^{-1}\left\{\frac{2r\pi}{T/\tau}\right\} = -\tan^{-1}\{\tau\omega\}$$

where we assume $(r\pi)^2 << T/\tau$. Once again we have seen why at small integration steps we have a fairly free choice of integration methods.

Of more importance, especially to the present discussion, is what happens at larger sampling ratios. Consider the specific value of $T/\tau = 0.3$. Then,

for various values of r, we obtain the data shown in Table 6 and plotted in Figure 27.

Table 6: Frequency-Response Comparison

Sampling Ratio (r)	Continuous System Gain (M)	Mixed-Mode/ True-Euler Gain (M_{mm}=M_{te})	Continuous System Phase (ϕ)	Mixed-Mode Phase (ϕ_{mm})	True-Euler Phase (ϕ_{te})
.001	.9998	.9998	- 1.200	- 0.8399	- 1.200
.003	.9980	.9986	- 3.595	- 2.517	- 3.597
.006	.9922	.9945	- 7.162	- 5.018	- 7.178
.010	.9788	.9850	-11.83	- 8.298	- 11.90
.030	.8467	.8854	-32.14	-22.78	- 33.58
.060	.6227	.6913	-51.49	-36.43	- 58.03
.100	.4309	.5018	-64.48	-43.49	- 79.49
.300	.1572	.2164	-80.96	-28.69	-110.4
.500	.0951	.1765	-84.55	0.0	-180.0

Note that these data do not show what happens as we increase T/τ. We already know that instability arises for $T/\tau > 2$.

The difference in performance to be expected from the two codes with $T/\tau = 0.3$ is clear. While both have less of a drop off in gain beyond the break frequency ($\omega = 1/\tau$, $r = T/\tau/2\pi$) than does the continuous system, their phase characteristics are dramatically different. The true Euler implementation has far more phase lag, and, if part of a larger simulation effort, could more readily induce overall system instability. The mixed-mode Euler has even less phase lag than the continuous system.

6.4 The Conclusion

The conclusion to be drawn from this analysis is obvious. One must not simply program an algorithm on the digital computer without regard for the effects of the manner in which the program is implemented.

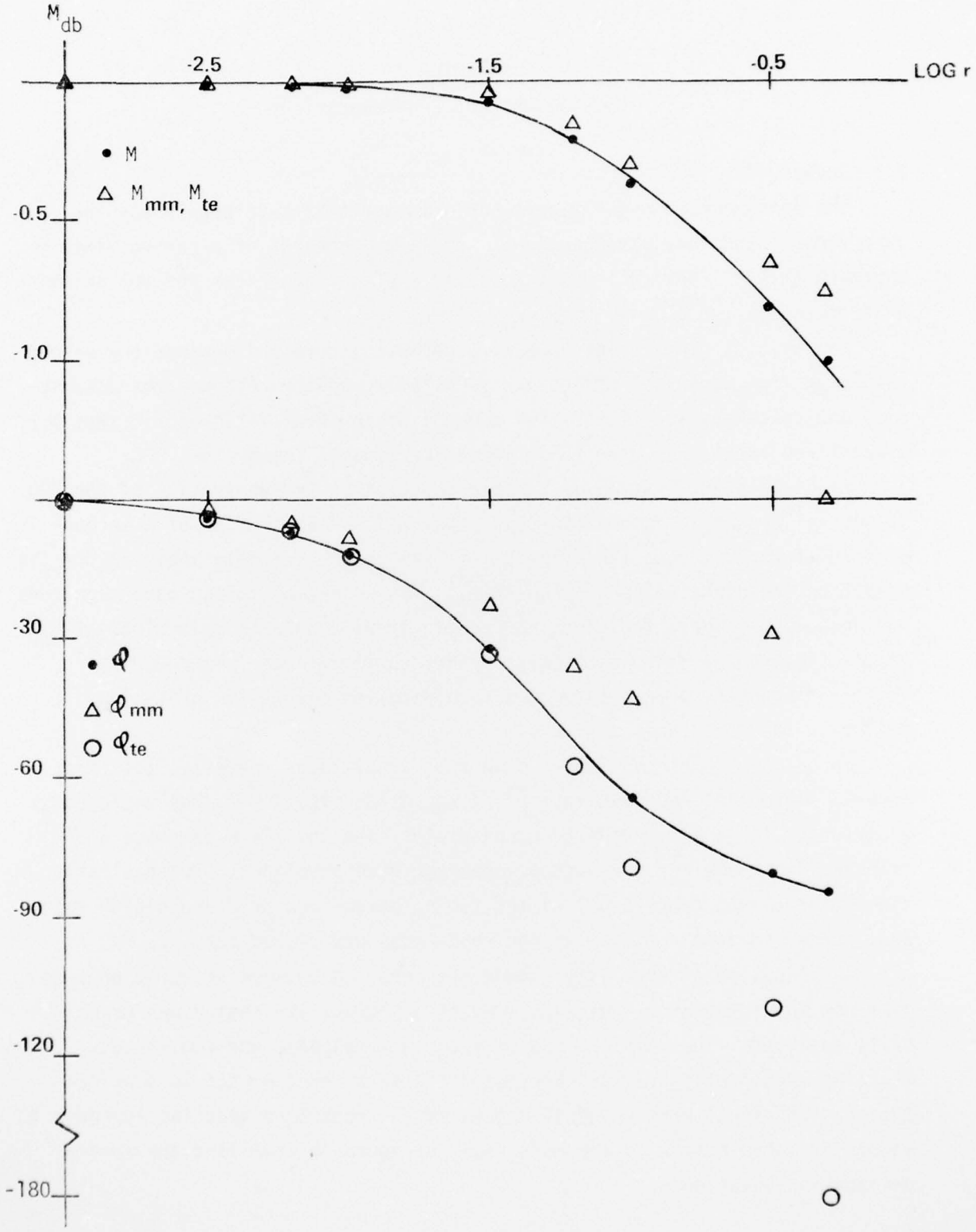


Figure 27. Comparison of Frequency-Response Characteristics

CHAPTER 7
Conclusions and Recommendations

7.1 Summary

The fundamentals underlying tunable integration have been presented in this report, and they have been analyzed in the context of a control-theory approach that is applicable to the analysis of any technique for the solution of differential equations.

The initial discussions were of a general nature and defined the errors resulting from numerical methods in terms of frequency-response characteristics and pole-zero locations. The concept of an ideal integrator, exact for a specified input, was used to motivate the tunable integrator.

A detailed explanation of tunable integration and derivation of the ZOH-TI was followed by a demonstration of tuning for the solution of a second-order differential equation. The ZOH-TI was then thoroughly analyzed for its open-loop frequency-response properties. The effect of tuning upon magnitude and phase-angle characteristics was cogently portrayed in Bode plots of M_{db} and ϕ. The use of these data for a priori tuning of the integrator was briefly discussed, preparatory to a more detailed discussion of tuning following the root-locus analysis.

To define the closed-loop system for a root-locus analysis, the first order differential equation $\tau\dot{x} + x = f$ was used. The ZOH-TI was employed as a corrector for a mixed-mode Euler predictor, and the block diagrams and transfer functions for the software package were developed. These clearly showed the effect (and value) of the tuning parameters on forward-path gain and, hence, on pole position as the root-locus was traced out.

A discussion of stability showed the coupling between integration step size and the integrator tuning parameters. A stability region was graphically depicted. The combined use of frequency-response and root-locus-stability considerations was then demonstrated as an approach to a priori tuning. The final portion of this part of the report related the approach to stability taken herein to the more familiar approach involving the error-propagation equation.

The final topic dicsussed was the effect of program implementation. It was demonstrated how an apparently innocuous change in program coding could seriously affect the frequency response of the integrator.

7.2 Conclusions

There are numerous conclusions to be drawn from the research reported on herein. I briefly state the more important ones below.

1. Control-theory methods of analysis are important and useful for the analysis of numerical methods.

2. The tunable integrator can be formulated as a low-order integrator with an inherent capability to adjust its frequency-response and pole-zero characteristics. This provides means other than variation of order or step size to control integrator performance. The result promises significant savings in computational burden and time.

3. Frequency-response and root-locus analyses provide the data for initial, albeit elementary, a priori tuning of the tunable integrator.

4. The method of implementing a chosen algorithm in computer code has a significant effect upon software performance. Control theory provides an effective way to analyze the potential performance of alternative code forms.

5. This report has not, nor was it intended to, developed a software package that in any sense can be thought of as a user-oriented tool in the manner of packages such as GEAR.

7.3 Recommendations

Again, simply enumerating, I make the following recommendations for follow-on work in this area.

1. Proposals have appeared in the literature for standardization of test problems used in evaluating numerical integrators.[17,18] Test cases such as those proposed by Krogh and others employed in [2] and [19] should be employed to thoroughly evaluate the performance range and error characteristics of the tunable integrators.

2. Many classical integrators can be shown to be special cases of the tunable integrators. The theory should be fully developed to tie the classical and tunable concepts together, to analyze the truncation model of error of

the former versus the phase-and-gain model of error of the latter, and to fully explain the behavior of tunable integrators. A start toward this end has been made in the control-theory analysis of this report.

3. At present, employing a tunable integrator on a complex system requires empirical determination of the proper tuning-parameter values. An elementary approach to a priori tuning has been discussed. Far more sophisticated and reliable approaches are required.

4. The Air Force and other technological communities have a potentially valuable tool at hand. The capability of tunable integration to solve problems where storage space and computation time are limited, but a degree of accuracy is required, must be experimentally determined by actual implementation whenever feasible.

BIBLIOGRAPHY

1. Hamming, R.W., Numerical Methods for Scientists and Engineers, second edition, New York: McGraw-Hill Book Company, Inc., 1973, Chapter 1 and Chapters 21-24.

2. Shampine, L.F., et.al., "Solving Nonstiff Ordinary Differential Equations - The State of the Art," SIAM Review, 18 (1976) 376-411.

3. Duke, A.A., et.al., "A Ballistic Trajectory Algorithm for Digital Airborne Fire Control," NWC-TP-5416, Naval Weapons Center, California, September 1972.

4. Smith, J.M., "Technology Seminar on Recent Developments in Modern Numerical Methods," Software Research Corporation, St. Louis, Missouri, 1973.

5. Smith, J.M., "Recent Developments in Numerical Integration," Journal of Dynamic Systems, Measurement and Control, 96 (1974) 61-70.

6. Sabin, M.L., "Bode Magnitude and Phase-Angle Characteristics of the Tunable Integrators," Modeling and Simulation, Vol. 6, Part 1, Proceedings of the Sixth Annual Conference, Pittsburgh: Instrument Society of America, 1975, pp. 505-509.

7. Sabin, M.L., "Tuneable Integration for Air Force Applications," Proceedings of the Symposium on Air Force Applications of Modern Control Theory, ARL 74-1060, Wright-Patterson AFB, Ohio, July 1974, pp. 221-236.

8. Smith, J.M., "Zero-Order T-Integration and its Relation to the Mean Value Theorem," Modeling and Simulation, Vol. 6, Part 1, Proceedings of the Sixth Annual Conference, Pittsburgh: Instrument Society of America, 1975, pp. 499-504.

9. Scheid, F., Numerical Analysis, Schaum's Outline Series, New York: McGraw-Hill Book Company, Inc., 1968, Chapters 6 and 14.

10. Hildebrand, F.B., Introduction to Numerical Analysis, New York: McGraw-Hill Book Company, 1956, Chapter 6.

11. Ralston, A., "Numerical Integration Methods for the Solution of Ordinary Differential Equations," Mathematical Methods for Digital Computers, Vol. 1, Ralston and Wilf, eds., New York: John Wiley and Sons, Inc., 1960, pp. 95-109.

12. Rosko, J.S., Digital Simulation of Physical Systems, Reading: Addison-Wesley Publishing Company, 1972, Chapters 5 and 6.

13. Tou, J.T., Digital and Sampled-Data Control Systems, New York: McGraw-Hill Book Company, Inc., 1959, pp. 198-207.

14. Fowler, M.E., "A New Numerical Method for Simulation," Simulation, J. McLeod, ed., New York: McGraw-Hill Book Company, 1968, pp. 275-281.

15. Smith, J.M., "Applications Memos - Modern Numerical Methods," Software Research Corporation, St. Louis, 1973.

16. Truxal, J.G., Control System Synthesis, New York: McGraw-Hill Book Company, 1955, Chapter 9.

17. Krogh, F.T., "On Testing a Subroutine for the Numerical Integration of Ordinary Differential Equations," Journal of the Association for Computing Machinery, 20 (1973) 545-562.

18. Krogh, F.T., "Opinions of Matters Connected With the Evaluation of Programs and Methods for Integrating Ordinary Differential Equations," Signum Newsletter, 7 (1972) 27-48.

19. Hull, T.E., et.al., "Comparing Numerical Methods for Ordinary Differential Equations," SIAM Journal on Numerical Analysis, 9 (1972) 603-637.